

The use of proteomic techniques to explore the holistic effects of nutrients *in vivo*

Helen R. Griffiths^{1*} and Melissa M. Grant²

¹*Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK*

²*School of Dentistry, Birmingham University, St Chad's Queensway, Birmingham B4 6NN, UK*

The availability of 'omics' technologies is transforming scientific approaches to physiological problems from a reductionist viewpoint to that of a holistic viewpoint. This is of profound importance in nutrition, since the integration of multiple systems at the level of gene expression on the synthetic side through to metabolic enzyme activity on the degradative side combine to govern nutrient availability to tissues. Protein activity is central to the process of nutrition from the initial absorption of nutrients via uptake carriers in the gut, through to distribution and transport in the blood, metabolism by degradative enzymes in tissues and excretion through renal tubule exchange proteins. Therefore, the global profiling of the proteome, defined as the entire protein complement of the genome expressed in a particular cell or organ, or in plasma or serum at a particular time, offers the potential for identification of important biomarkers of nutritional state that respond to alterations in diet. The present review considers the published evidence of nutritional modulation of the proteome *in vivo* which has expanded exponentially over the last 3 years. It highlights some of the challenges faced by researchers using proteomic approaches to understand the interactions of diet with genomic and metabolic–phenotypic variables in normal populations.

Proteomic techniques: Nutritional biomarkers: Proteins: Analytical methods

Introduction

Technologies have advanced significantly in the past 5 years to enable us to ask holistic questions such as 'what proteins/pathways does glucose modulate (for example, through post-translational modification, altered compartmentalisation or expression/degradation) in islet cells?' rather than the traditional reductionist approach to address an hypothesis in a selected target tissue. Embracing this paradigm shift, the open-ended approach to investigating mechanisms of physiological effects can lead to the identification of novel regulatory pathways possibly leading to new roadmaps for research. The value of findings derived from 'fishing' approaches is subsequently validated using conventional biochemical, cellular and molecular biological approaches both *in vitro* and *in vivo*. Thus, scientific rigour is maintained and novel mechanistic influences on physiology can be identified.

The systems biology approach offers significant strengths to the field of nutrition, where the influence of both genotype and phenotype and external influences have marked effects on absorption, distribution, metabolism and excretion and

therefore the nourishment of an individual organism. Whilst systems biology integrates the temporal study of the transcriptome (all mRNA molecules (or transcripts) in one cell or tissue), the proteome (total protein complement of a genome present in cells and/or tissue at a given time) and the metabolome (entire complement of all the small-molecular-weight metabolites inside a cell suspension or organism), there can be little correlation between these approaches – not all changes in mRNA expression are translated directly to changes in protein levels. The correlation coefficient in a controlled experimental unicellular environment rarely exceeds 0.5 and many changes in protein levels can be attributed to changes in mRNA stability; mRNA is around for a longer or shorter period of time before degradation and therefore is more or less likely to be translated to protein. Similarly, an expressed protein is not necessarily an activated protein and therefore may not contribute to the metabolome. Such verification is an important qualifier for emergent findings from systems biology studies.

Whilst recognising the need for experimental confirmation of any change in protein level on functional effect,

Abbreviations: CLA, conjugated linoleic acid; 2DE, two-dimensional electrophoresis; DIGE, difference in gel electrophoresis; ICAT, isotope-coded affinity tag; MALDI, matrix-assisted laser desorption ionisation; SELDI, surface-enhanced laser desorption ionisation; TOF, time of flight.

* **Corresponding author:** Professor Helen R. Griffiths, fax +44 121 359 5142, email h.r.griffiths@aston.ac.uk

the application of proteomics can be easily rationalised in nutrition research; proteins are ultimately responsible for the absorption and distribution of nutrients (glucose and amino acid transporters, apolipoproteins), metabolism (proteases, lipases, glycolytic enzymes) and excretion (renal transporters). At the time of preparing the present article, over the past 5 years, a PubMed search using the medical subject heading (MeSH) terms nutrition AND proteomics revealed thirty-one published review articles. Of note, these articles focused on the potential for a revolution in our understanding of nutrition through application of proteomic techniques, and, as is often the case for emergent technologies, exceed the total number of studies currently published which have used proteomics to analyse nutritional effects *in vivo*. In order to consider the likely future value of proteomics in nutrition research, the present review has three objectives: (1) to summarise current methodological approaches; (2) to highlight the application of proteomics to the study of nutrition *in vivo*; (3) to consider any possible limitations which may currently hold back widespread application of proteomics and the future of proteomics in nutrition research.

Current methodological approaches

The publication of the human genome led to the realisation that many of the mechanisms underlying disease and influencing physiology are expressed at the supra-genomic level, and this had a catalytic effect on the development of proteomic methods. Since the 1970s, the traditional workhorse of high-resolution protein separation has been two-dimensional electrophoresis (2DE) (Anderson & Anderson, 1996). However, it was not until the late 1990s that this tool was widely applied. This revolution was caused by advances in MS, particularly in easy-to-use mass spectrometric techniques that became readily accessible to life scientists. More recently, other technologies have been developed which offer increased speed through a more simplistic approach. What follows is a brief overview of proteomic technologies. For more detailed consideration of the methods and their applications, the reader is referred to Aldred *et al.* (2004).

All the technologies used in proteomics are held to ransom by pre-analytical variation incurred in sample collection, handling and storage, particularly of clinical samples where there may be a considerable amount of time between collection and analysis. Nevertheless, a recent study demonstrated that length of time to freezing (up to 24 h) does not significantly affect surface-enhanced laser desorption ionisation (SELDI)–time of flight (TOF) MS variability in blood or urine (Traum *et al.* 2006).

Sample preparation

For all proteomic techniques, good sample preparation is the key to success. The composition of any particular sample will differ according to source; however, most samples will encounter problems associated with the simultaneous presence of highly abundant proteins and low copy number proteins causing a sensitivity problem over the wide dynamic range in protein quantification. If plasma

is used as an example, the dynamic range in protein concentration is about six orders of magnitude. The most abundant protein, albumin, and the following nine most abundant proteins make up nearly 70% of the plasma protein composition. Thus, any scarce proteins are masked, hard to find and difficult to observe for changes in expression or modification. A compounding feature of the plasma matrix is that the low-abundance proteins are also possibly the most interesting; tissue exudates, cytokines and growth factors may all lie hidden. Therefore, the first step in sample preparation can be their simplification by lowering the complexity of the sample; this can be achieved either by removing interfering substances or enriching in certain species of interest (Righetti *et al.* 2005). For example, high-abundance proteins such as albumin and globulins can be removed by immunoprecipitation techniques, although it should be noted that whilst removing high-abundance proteins, other associated proteins may also be extracted, highlighting the need for verification of findings using non-manipulated samples. Alternatively, specific pools of proteins can be analysed. These include: proteins with a particular post-translational modification such as phosphorylation which can be immunoprecipitated (Kabuyama *et al.* 2004); proteins in particular subcellular locations such as the mitochondrion which can be enriched using ultracentrifugation techniques (Warnock *et al.* 2004); or newly transcribed proteins which can be metabolically labelled.

Gel-based technologies

The most widely used of the gel-based proteomic technologies is 2DE utilising isoelectric focusing and SDS-PAGE, where proteins are separated by charge in the first dimension and mass in the second. The technological advancement in isoelectric focusing has enabled single pI unit range immobilised pH gradient strips, zoom strips, which allow for very high-resolution separations. Following 2DE, each gel comprises a series of spots where each spot corresponds to one or more protein species. Alternatively, a subset of proteins can be isolated by using Western blotting for post-translational modifications to further limit the number of proteins taken through to analysis. Quantitative protein expression profiling is achieved by visualising gels post-staining with one of a variety of chromophores such as Coomassie Blue or silver (both of which have a limited dynamic range), and more recently fluorescent dyes such as the Sypro family have improved the sensitivity and dynamic ranges.

The fluorescent cyanine dyes (Cy2, Cy3 and Cy5), which carry n-hydroxysuccinimidyl ester functionality and bind the ϵ -amino group on lysine residues, have also been introduced to 2DE. This allows the labelling of three spectrally resolvable fluorophores and enables the simultaneous visualisation of test, control and standard gels (Kolkman *et al.* 2005). The difference in gel electrophoresis (DIGE) approach has been validated for use in quantitative proteomics although cost precludes its routine application.

Once visualised, the gels' images are captured and *in silico* representations are analysed by a variety of algorithms and software options. Proteins of interest, once

identified, are then excised from the gel or membrane and subjected to limited trypsinisation. The resulting peptides are extracted and identified using either matrix-assisted laser desorption ionisation (MALDI)–TOF MS or tandem MS techniques (see later). Another emerging technology is the application of column separation by capillary electrophoresis (Fliser *et al.* 2005), which has the advantage of being able to cope with crude samples with high salt concentration such as urine. Peak-picking algorithms are used to differentiate the peptide between control and test before sample identification by MS.

Chromatography-based technologies

Chromatographic methods based on different matrices have been used to separate proteins for over 50 years. For many proteomic applications, proteins in a sample are first digested to peptides and then these are separated chromatographically before mass spectral protein identification. The simplest methods merely slow down the rate at which the peptides enter the analyser so that the maximum amount of time can be taken to identify as many of the peptides as possible, so called shotgun proteomics. Multi-dimensional chromatography techniques, such as multi-dimensional protein identification technology which uses columns consisting of strong cation exchange material back-to-back with reversed-phase material, separate peptides before elution and ionisation within the mass spectrometer, in much the same manner (Wolters *et al.* 2001). However, many techniques are used to extract or highlight peptides of choice. For example, immobilised metal ion affinity chromatography utilises chromatographic media with bound Ni ions to extract phosphopeptides from complex mixtures, thus enabling the identification of peptides from proteins involved in signalling cascades (Riggs *et al.* 2001). Isotope-coded affinity tag (ICAT) methodologies use stable-isotope tags to differentially label two samples which can then be mixed. The mixture is then chromatographically separated and quantitatively identified by MS (Gygi *et al.* 1999, 2002). The most commonly used tags bind to cysteine residues in the proteins; thus fewer peptides are labelled. Stable-isotope labelling with amino acids in cell culture (Ong *et al.* 2002), a complementary technique to ICAT, labels the proteins with light or heavy C (^{12}C or ^{13}C) or N (^{14}N or ^{15}N) usually incorporated into the samples during growth in medium containing labelled lysine. The resulting processed samples are then analysed by MS and give quantitative answers as with ICAT. However, this does require the synthesis of new proteins and thus is only of use in cell-culture work or in fast-growing organisms as its name suggests.

Protein arrays

The success of gene arrays for transcriptomics has led to the idea of protein arrays. However, with nothing analogous to PCR for the amplification of signal and the lack of a single complimentary hybridisation substrate, the development of the technology has proven to be more difficult. Different interacting molecules such as antibodies, DNA and drug molecules have been used as bait to identify

components of protein samples. Once the proteins have bound to the bait they must then be identified and quantified, requiring the application of MS (Pavlickova *et al.* 2004). In fact, this baiting approach is a natural development from the basis of SELDI–TOF MS (Issaq *et al.* 2002). The SELDI approach was pioneered by Ciphergen to enable the analysis of complex proteomes such as plasma and enables the selective capture of proteins by varying the nature of the capture matrix.

Protein identification

All of the above techniques are reliant on the successful identification of the protein or peptide. The favoured technique for this is MS (Yates, 2004; Barrett *et al.* 2005). The two most commonly used technologies are MALDI–TOF MS and electrospray ionisation MS. The first uses solid-state peptides and the latter liquid-state peptides. MALDI–TOF MS has become the choice for many life scientists because of the ease of use and of analysis. Essentially, it produces mass-mapping of an unknown single protein derived from a spot on 2DE gel by breaking the protein into specific peptide fragments using trypsin which cleaves after lysine or arginine residues. Mass spectrometric analysis of the enzymic digest generates a ‘mass-map’ which is unique to the digested protein and allows unambiguous identification of mass fingerprints through public databases using the Mascot search engine. The probability of protein identifications being correct is calculated according to the ‘MOWSE’ score (molecular weight search based on peptide scoring frequency) defined by Pappin *et al.* (1993).

Electrospray ionisation MS lends itself more easily to tandem techniques where peptides can be sequenced directly and thus provides greater security to protein identification, over mass fingerprinting. Once a protein is identified, it is essential to ensure that the proteins or peptides identified are contextually accurate. This can often be achieved by a complementary technique such as Western blotting or fluorescence microscopy, where the lack of need for sample manipulation should add strength to the observation.

Application of proteomics to nutrition

Whilst the number of articles describing proteomic applications to nutrition has increased exponentially since 2002, the majority of these approaches describe single-cell models in tissue culture. This can provide some important evidence for mechanisms of nutrient effect but can provide limited information on systems interactions. For that reason, the following section is limited to a review of those studies undertaken *in vivo*, both in human and in animal models. Application of the *in vivo* limit to PubMed searches led to the recovery of four major experimental themes: those investigating the pathophysiology of dietary deficiency states; those investigating the mechanisms by which nutrients may modify disease states; those evaluating the effects of supplementation with micro- and macronutrients and excess of particular nutrients on otherwise normal subjects or animals. In many respects it is too early to consider concordance between nutrition–proteomic studies

as no two studies are the same; however, these articles are considered in a systematic way and attention is paid to whether the articles have verified their findings using alternative experimental approaches.

Use of proteomics to investigate mechanisms underlying pathophysiology of nutritional deficiency states

For ethical reasons, the study of experimentally induced nutritional deficiency states is undertaken in animal models; however, in addition to understanding the multiplicity of pathophysiological states, one of the possible outcomes of such studies is the identification of possible early biomarkers of functional deficiency which may be applied to vulnerable populations (Go *et al.* 2005).

The consequences of prolonged K⁺ deficiency on the mouse renal proteome have been investigated by Thongboonkerd *et al.* (2006) as hypokalaemia is known to be associated with a number of complications including metabolic alkalosis, polyuria and renal tubule injury. These workers applied 2DE differential analysis of the kidney proteomes from normal or dietary depleted mice (8 weeks) and reported thirty-three differentially expressed proteins in the kidney proteome of K-deficient mice. Using an MS-based approach, thirty of these proteins were identified and fell into three major protein function classes; metabolic enzymes, signalling proteins and cytoskeletal proteins. Some of the metabolic enzymes identified are involved in metabolic alkalosis and thus warrant further study as putative effectors of K⁺ deficiency.

A consideration of the subclinical effects of Cu and Fe deficiency has been recently reported by Tosco *et al.* (2005). These metals are important cofactors for a number of enzymes involved in normal physiological processes and their deficiency has widespread consequences in brain development and vascular function (Schuschke, 1997; Prohaska, 2000). Differential analysis of the rat intestinal proteome showed significant changes in the expression of proteins associated with glucose and fatty acid metabolism, molecular chaperones, cytoskeleton and vitamin transporters. Again, this work highlights novel pathways for considering the consequence of unbalanced micronutrient intake, with alterations to the cytoskeletal proteome being modified in both studies.

Chanson *et al.* (2005) investigated the effects of folate deficiency on the rat liver proteome as low liver folate has been shown to increase abnormal one-C metabolism and the risk of degenerative disease. Young rats were fed an amino acid-defined, folate-deficient diet for 4 weeks and compared with animals receiving a normal diet. Again, a 2DE approach was adopted and differentially expressed proteins were determined by MALDI-MS. These workers identified five up regulated proteins and four down regulated proteins in the folate-deficient rats; the up regulated proteins included glutathione peroxidase and peroxiredoxins which are indicative of an increased level of hepatic oxidative stress (Cessarato *et al.* 2005). Importantly, these researchers confirmed the observation that glutathione peroxidase levels are up regulated in the liver of folate-deficient animals using a Western blotting approach.

Using proteomics to investigate mechanisms of disease process using nutrients and micronutrients

Certain micronutrients, including Se and tocopherol, have been suggested to have a chemopreventive effect in prostate cancer (Rayman, 2005). To investigate the potential for defining a disease phenotype which may be modified by micronutrient supplementation, Kim *et al.* (2005) have studied the plasma proteome of controls and prostate cancer patients after dietary supplementation with vitamin E, Se, both or placebo for 3–6 weeks. Pre-fractionation of plasma to examine the lower-molecular-weight (2–13.5 kDa) proteins by using SELDI-MS was adopted as the method of choice and, using principal component analysis, these workers were able to differentiate a prostate cancer plasma proteome from the matched control. Moreover, they showed that the combination of Se and vitamin E induced significant changes in the cancer patients' proteomes towards a profile indicative of prostate-cancer-free status. Identification of a disease *v.* normal protein profile provides an approach which can be used to monitor therapeutic responsiveness without any need for identification of the proteins involved.

A similar approach, using proteomics to identify specific multi-protein profiles associated with pathology which may not be evident from conventional methods, has been undertaken by Weissinger *et al.* (2006). These workers adopted a shotgun proteomics approach to investigate the effects of oral vitamin C supplementation in haemodialysis patients. They were able to define a plasma polypeptide fingerprint comprising thirty different species that characterised patients with renal dysfunction undergoing haemodialysis and, furthermore, were able to show that several of the polypeptides were normalised following supplementation with vitamin C (250 mg/d) for 3 weeks. Whilst definition of these polypeptides is of interest to provide insight into the functions of vitamin C in haemodialysis patients, it is not necessary to know their identity, only their *m/z* signal, to follow the response to diet or treatment.

Using a mouse model of mammary cancer (with dimethylbenz[a]anthracene as the tumour promoter), Rowell *et al.* (2005) adopted a conventional proteomic approach based on 2DE and MALDI-TOF analysis to define a novel mechanism of chemoprotection afforded by the soya isoflavone, genistein. Whilst six proteins appeared to be altered by pre-pubertal exposure to genistein but not daidzein (500 µg/g), five were further investigated and up regulation of only one, GTP-cyclohydrolase-1, was confirmed by Western blotting. To pursue the importance of this finding further, the authors considered the downstream signalling from GTP-cyclohydrolase-1 and were able to confirm that tyrosine hydroxylase was up regulated whereas vascular endothelial growth factor receptor 2 was down regulated in rats at 50 d, when administered with genistein in the pre-pubertal period. Thus, a proteomic approach allowed the postulation of a novel mechanism for the chemoprotective effects of genistein through inhibition of angiogenesis.

A similar approach was adopted by Poon *et al.* (2005) to further investigate the observation that α-lipoic acid can reverse memory impairment in the senescence-accelerated-prone mouse (strain 8; SAMP8). Using 2DE separation with

LC-MS/MS to define proteins with altered expression profiles, these authors reported that lipoic acid supplementation was able to prevent the development of the accelerated ageing phenotype. Specifically, they observed restoration of expression of neurofilament triplet L protein, α -enolase and ubiquitous mitochondrial creatinine kinase to levels seen in normal mice and prevention of the oxidation of lactate dehydrogenase B, dehydropyrimidinase-like protein 2 and α -enolase in mice receiving α -lipoic acid. However, further functional studies or application of other techniques to confirm findings are not described in this article.

There are many studies suggesting that dietary fatty acids, particularly fish oils, may have a significant effect on atherosclerosis, but a recent meta-analysis has not supported this finding (Hooper *et al.* 2006). Given this disparity in the literature, further evaluation of the effects of such fatty acids on vascular function is warranted. In this regard, in the medium term, proteomics may offer some insight into the variable outcomes of intervention studies. In the short term, investigation of the effects of dietary fatty acids in animal models, specifically effects of fish oil, *cis*-9, *trans*-11- and *trans*-10, *cis*-12-conjugated linoleic acid (CLA) or elaidic acid on the liver proteomes of mice with hypercholesterolaemia due to apo E deletion or expression of the apo E* Leiden transgenic mice (de Roos *et al.* 2005a,b) may also provide some clues. The first study (de Roos *et al.* 2005a) used a conventional 2DE approach to examine differential protein expression and principal component analysis to determine which proteins were sensitive to the effects of fish oil. This investigation showed that fish oil had a major effect on cytosolic proteins but that the effects of elaidic acid were restricted to membrane proteins. In addition, the authors were able to correlate physiological effects on plasma insulin, glucose, cholesterol and fatty acid levels with proteomic changes; a change was observed in levels of the proteins, long-chain acyl-CoA thioester hydrolase and adipophilin, which contributed to the induction of a phenotype consistent with the metabolic syndrome after dietary enrichment with CLA. The more recent study focused on supplementing with dietary CLA in apo E knockout mice (de Roos *et al.* 2005b); the principal defining protein expression change following *cis*-9, *trans*-11-CLA administration was the up regulation of hsp70 family members. The differential effect of *trans*-10, *cis*-12-CLA was confirmed by principal component analysis; the key enzymes up regulated were enzymes associated with gluconeogenic, β oxidation and ketogenic pathways. In this model, increased levels of hepatic serotransferrin, an acute-phase reactant, were associated with a phenotype of insulin resistance, thus making a tantalising link between the metabolic syndrome and inflammation which merits further study.

Effects of dietary supplementation or enrichment on the proteomes of healthy organisms

The studies described earlier give some indication that disease activity or outcome may be modulated by nutritional means; however, in the mammary cancer model, dietary intervention was required before induction of the cancer.

Similarly, vascular occlusion in subjects presenting with clinical manifestations of disease is likely to have developed over three decades or more and the potential for dietary supplements to reverse such changes seems remote. As epidemiological evidence is in support of a beneficial effect for many micronutrients on health outcomes this may be due to early intervention before disease can become established, thus the advancements offered by proteomic technologies may facilitate an understanding of preventative mechanisms in healthy subjects. The following five articles detail proteomic investigations of nutrient effect in normal subjects and animals.

Our recent paper (Aldred *et al.* 2006) describes the use of 2DE with MALDI-MS to investigate the plasma proteome response to increasing doses of α -tocopherol over 4 weeks. Subsequent analysis of the pooled 2DE proteomes led to the identification of increased expression of pro-apo A1 following tocopherol supplementation, with time- and dose-dependent effects. This was confirmed in each individual non-manipulated plasma sample to ascertain that the effect was not due to albumin depletion or due to pooling. Importantly, this work demonstrated a novel mechanism of benefit in healthy subjects, that tocopherol may increase hepatic synthesis of apo A1 which is involved in reverse cholesterol transport and is inversely related to CVD risk.

The effect of supplementation of the diet with cruciform vegetables on the serum proteome of healthy subjects was reported recently by Mitchell *et al.* (2005). These workers applied MALDI-TOF MS to serum samples depleted of major proteins such as albumin, and applied logistic regression models and peak-picking algorithms to distinguish participants who had followed a 7 d diet with cruciform vegetables from those following a 7 d diet without the vegetables. The technique proved powerful enough to classify the participants' diets to 76% accuracy using two *m/z* peaks; one of these peaks was subsequently identified as the B-chain of α 2-HS glycoprotein, a serum protein previously found to be involved in insulin resistance (Dahlman *et al.* 2004).

There is increasing evidence for the benefits of anthocyanadins against cognitive decline, although the mechanisms underlying these effects are unknown (Galli *et al.* 2002). To investigate this further, Kim *et al.* (2006) have undertaken a proteomic investigation into the potential physiological benefits of 6 weeks of grape-derived dietary supplements on the whole rat brain. Using a 2DE approach and peptide mass fingerprinting by MALDI-MS, thirteen proteins were determined to be different in the brains of animals exposed to the grape supplement compared with control brains. These included proteins associated with energy generation, such as creatine kinase, heat-shock proteins and cytoskeletal proteins including neurofilament protein light chain. The latter was described by Poon *et al.* (2005) as being restored in the brains of accelerated ageing mice following administration of lipoic acid, again supporting the hypothesis that neurofilament protein may have an important role in preventing cognitive decline.

The final studies which meet the criteria for inclusion as proteomic studies detailing the effects of nutrients on normal organisms describe different objectives; to evaluate whether diet improves meat quality through reducing

oxidative degradation of muscle proteins or improves growth rate. The first of these by Stagsted *et al.* (2004) used a proteomic approach to separate chicken muscle proteins together with an immunoblotting approach to detect oxidation as protein carbonyls or nitrotyrosine. In common with the earlier report of Poon *et al.* (2005), α -enolase was found to be highly susceptible to oxidation and in this study, the levels of oxidation of enolase in chicken muscle were reduced in chickens fed a diet supplemented with antioxidant-rich fruit and vegetables compared with those animals receiving a low-antioxidant diet.

The second related study, undertaken by Martin *et al.* (2003) in rainbow trout, aimed to determine whether substitution of fish meal for soya meal had an effect on growth and, if so, to discover the mechanism using a proteomic approach. Using a 2DE approach with MALDI-TOF MS for peptide mass-mapping, thirty-three peptides were differentially expressed between control and soya-fed animals but no difference in growth rate was reported between the diets. The change in expression of transcripts for two proteins, apo A1 and aldolase B confirmed the proteomic findings and suggested that fish fed the soya diet were exhibiting increased metabolism of proteins and cholesterol. In support of our work in healthy human subjects (Aldred *et al.* 2006) apo A1 expression has again been shown to be sensitive to dietary nutrients.

These latter two studies demonstrate the potential for commercial gain through application of proteomics to study the mechanisms of improved productivity.

Effects of dietary excess on the proteomes of otherwise healthy organisms

Whilst there are also reports in the literature which have adopted a proteomic approach to describe the effects of nutrient excess on physiology, they are beyond the scope of the present review. The findings from these studies are summarised in Table 1 and the reader is referred to the original publications for further information.

Many of the articles reviewed earlier have focused on the potential for nutrients to increase protein expression. However, the steady-state level of a particular protein species is the summation of both synthetic and degradative changes. There are little data available yet on how nutrients

may regulate turnover. However, a proteomic approach has led to the postulate that post-translational modification of proteins by *o*-GlcNAc may occur when cellular glucose concentrations rise and this may inhibit proteasomal degradation of transcription factors and therefore allow cells to regulate transcriptional activity according to nutritional status (Zachara & Hart, 2004). When they have an energy source, they can transcribe proteins and in the absence of glucose, proteasomal degradation of transcription factors is enabled.

Current limitations

Whilst these studies are beginning to demonstrate the value of proteomics, any summation of data can only be considered valid if there is confidence in the analytical methodology *per se*. Several important caveats remain and require closer inspection:

How recoverable is the cellular or plasma proteome?

How accurate are the quantitative proteomic techniques?

What is the likelihood of statistical anomaly or false positives through multiple sampling error?

The current approaches to solving these methodological issues are described and will allow the power of proteomics to be confidently accepted and potential benefits for nutrition research to be recognised.

How recoverable is the cellular or plasma proteome?

As alluded to earlier in the present review, solubility problems can limit the representation of highly hydrophobic membrane proteins which are difficult to recover using commonly available detergents. New detergents are being developed to extend the application of proteomics to low-solubility proteins (Rabilloud, 2003; Stanley *et al.* 2003), in particular in the development of zwitterionic detergents such as C7BzO. However, there is no universal method for extracting different membrane proteins and individual optimisation is necessary using a panel of detergents.

A different challenge faces those studying the plasma proteome, namely the wide dynamic range of protein components which is in the order of 10^{12} . Lower-abundance proteins are masked by high-abundance proteins such as albumin and immunoglobulins, as previously described on

Table 1. Key features of extracted data from proteomic studies of 'overnutrition'

| Nutrient | Species | Tissue examined | Key observation | Reference |
|-------------|---------|--|---|------------------------------|
| Cholesterol | Mouse | Liver | Showed testosterone deficiency (phenotype); expression of androgen-regulated proteins carbonic anhydrase III and glutathione S-transferase P2 also reduced | Feng <i>et al.</i> (2005) |
| Cholesterol | Mouse | Blood vessels | Cys peroxiredoxin was oxidised and correlated with extent of lesion formation | Mayr <i>et al.</i> (2005) |
| Fat | Mouse | Gastrocnemius muscle, white and brown adipose tissue and liver | Brown adipose tissue increased expression of tricarboxylic acid cycle enzymes and respiratory chain proteins, suggesting increased energy expenditure to defend against weight gain | Schmid <i>et al.</i> (2004) |
| Cu | Sheep | Liver | Increase in cytosolic isocitrate dehydrogenase, mitochondrial thioredoxin-dependent peroxidase reductase and cathepsin D in liver which associated with acute mitochondrial damage | Haywood <i>et al.</i> (2005) |

p. 285. Albumin can be 'selectively' removed using chemical (cibacron blue) or antibody techniques; however, albumin is tightly associated with many small peptides in plasma and the removal of albumin is also likely to deplete hormones including thyroid and steroid hormones. Initial fractionation using strong detergents before the removal of high-abundance proteins offers a more secure route for analysing low-abundance proteins without the problems of inadvertent peptide depletion; however, detergent denaturation also destroys the properties that allow certain components to be depleted. The recent Human Proteome Organization Plasma Proteome Project report (Omenn *et al.* 2005) has highlighted the key issues to be addressed over the next few years to improve the value of data which can be derived from plasma samples.

The SELDI-TOF MS approach has been applied to several of the studies reviewed in the present article, however, it preferentially detects low-molecular-weight peptides as they have a higher efficiency of ionisation. This technology also poses some challenges in protein identification which is based on peptide mass matching of complex mixtures and can not inherently accommodate post-translational modifications; the lack of identity of the putative biomarkers remains unaddressed. Overall, SELDI has been suggested to suffer with high noise levels, and from low sensitivity and specificity (Diamandis, 2004). Indeed, protein concentrations responsible for the 'peaks' in SELDI are normally in the $\mu\text{g/ml}$ range, much higher than many known biomarkers for cancer. The plasma proteome Human Proteome Organization report (Rai *et al.* 2005) has recommended stringent standardisation and pre-fractionation to increase the utility of the SELDI approach after intra-laboratory CV revealed variations from 15 to 43% across five laboratories.

How accurate are the quantitative proteomic techniques?

Stable-isotope labelling (ICAT) has been a core technology for use in quantitative proteomic MS; however, there can be some issues with variability due to incomplete labelling and the need for such experiments to be conducted in the presence of isotopes. Recent attention has focused on peptide peak area quantification and this technology has been successfully applied to characterisation of the proteomes of wild-type and p53-deficient HCT-116 human cells with acceptable reproducibility (Wang *et al.* 2006). Comparison of ion peak areas following analysis by ion trap or Fourier transform MS provided CV in the order of 10%. However, such an analysis will normally follow on from the identification of spots by 2DE which differ at least two-fold and this stage alone can introduce errors in spot identification through differences in sample loading between gels, gel-to-gel warping and variation during fixation. Whilst complex algorithms exist in 2DE analytical software packages to maximise intelligent matching, this is usually overseen by an experienced operator and as such presents a huge bottleneck in comparative proteomics. For this reason, the development of DIGE has afforded significant benefits; whilst expensive to use routinely, it offers the potential to directly compare two samples within the same gel. As with all 2DE analysis, some spots may not

be detected by DIGE if their isoelectric points or molecular weights lie outside the range of the immobilised pH gradient strips. Such a problem is not encountered in direct MS techniques.

In a recent study designed to examine the consistency between DIGE and ICAT, there was limited overlap between detected proteins, suggesting that such techniques should be regarded as complementary rather than alternative methods of analysis (Wu *et al.* 2006).

Size of effect and sample size in proteomics

From the earlier discussion, it is evident that the combination of experimental variation and intrinsic complexity of mixed protein samples contributes to the background 'noise' in proteomics. This so called 'high dimensionality' of the data, due to the equivalent or smaller number of samples being analysed compared with the number of data points or protein spots being detected, increases the risk of 'overfitting' the data. To overcome this problem and minimise the risk of identifying false positives, machine learning techniques have been developed to reduce the dimension of putative biomarker proteins and yield a manageable set of proteins for further validation. A recent report has described the application of support-vector machine learning techniques to simulated datasets to demonstrate the robustness against identification of outliers and subsequently to real SELDI datasets from a breast cancer proteomics study (Zhang *et al.* 2006). Biomarkers 'identified' by the algorithm were subsequently validated in biological experiments; for example, a peptide that was sequenced by a direct on chip sequencing technique was followed up in breast cancer patients and found to associate with disease status.

In addition to complexity in the technology which creates 'high-dimensional' datasets, the genetic variability in human studies adds further complications to the application of proteomics to nutritional intervention studies in human subjects. One of the key issues in considering nutrition in healthy subjects is that homeostatic mechanisms which serve to maintain normal physiology are, by definition, likely to reduce any effects of diet on the proteome. This is less of a problem in animal studies using in-bred species. However, in a cohort of human subjects recruited to a nutritional intervention study, inter-individual variability due to genotypic differences in any given proteome is likely to exceed the changes induced by environment or diet. In our recent study of the plasma proteome following α -tocopherol supplementation (Aldred *et al.* 2006), with ten or eleven subjects in each of three supplementation groups measured at baseline, after 2 weeks and after 4 weeks, the evidence suggested that inter-individual variation in protein expression exceeded that of the supplement. In order to circumvent these issues, samples from subjects in the same supplement group and at the same time point were pooled. Such pooling reduces individual noise between subjects but also may lead to the loss of individual responsiveness and possible dilution of effect; therefore the need to confirm 2DE findings using native, individual samples is very important.

Pooling has also been necessitated in the past when cohort sizes are large; for example, an experiment with

twelve different paradigms and three replicate gels per experiment or population-based studies with fifty individuals under two different regimens, also using replicate gels, the number of gels becomes costly, and the extent of analysis, unwieldy.

Current software uses algorithms to match gels, allowing for the analysis of protein quantity. The proteins of interest can then be found by using Boolean or fold changes. Advances in bioinformatics and multiple component analysis to enable analysis of large numbers of gel datasets (> 100), each with up to 1000 spots per gel, will provide a route to work through such inter-individual variability and large sample sizes to define consistent patterns of effect without the need for sample pooling.

Future approaches

The technology underlying proteomics continues to progress from the most basic level in improving dyes which can provide linearity with protein expression over a wide dynamic range, to improving software packages to deal with multiple gels and associated intelligent systems to analyse data with minimal bias. Many steps are being taken towards miniaturisation, both of sample size use and of instruments, because of the limited and unique samples which are available from clinical settings, including the moves towards lab-on-a-chip scenarios (Diks & Peppelenbosch, 2004; Marko-Varga *et al.* 2004). Advances in MS are making it possible to do simultaneous tissue imaging and profiling, putting biochemical findings into the context of disease state. Electron tomography is beginning to produce three-dimensional maps of cells and proteins (Baumeister, 2005).

Further steps beyond proteomics such as temporal and spatial analyses will help build systems biology information (Papin & Subramaniam, 2004). However, 70–80% of the proteome still remains to be seen and so new separation techniques, such as (1) multistacking chromatography which consists of a set of immobilised chemistries, serially connected in a stack format to resolve proteins before MS and (2) solid-phase ligand libraries where a library of combinatorial ligands is coupled to small beads before mixing with a complex proteome to significantly reduce the concentration differences in between proteins such as plasma, need to be adopted to increase visualisation of proteins (Righetti *et al.* 2005). The first steps towards sharing and standardisation of proteomic and mass spectral data has begun through projects such as 'Minimum Information About a Proteomics Experiment' (known as MIAPE) by the Human Proteomics Organization. This will facilitate data comparison between laboratories and will allow the importance of proteomics in nutrition to be fully realised.

References

- Aldred S, Grant MM & Griffiths HR (2004) The use of proteomics for the assessment of clinical samples in research. *Clinical Biochemistry* **37**, 943–952.
- Aldred S, Sozzi T, Mudway I, Grant MM, Neubert H, Kelly FJ & Griffiths HR (2006) α -Tocopherol supplementation elevates plasma apolipoprotein A1 isoforms in normal healthy subjects. *Proteomics* **6**, 1695–1703.
- Anderson NG & Anderson NL (1996) Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* **17**, 443–453.
- Barrett J, Brophy PM & Hamilton JV (2005) Analysing proteomic data. *International Journal for Parasitology* **35**, 543–553.
- Baumeister W (2005) From proteomic inventory to architecture. *FEBS Letters* **579**, 933–937.
- Cessarato L, Vascotto C, D'Ambrosio C, Scaloni A, Baccarani U, Paron I, Damante G, Calligaris S, Quadrifoglio F, Tiribelli C & Tell G (2005) Overoxidation of peroxiredoxins as an immediate and sensitive marker of oxidative stress in HepG2 cells and its application to the redox effects induced by ischemia/reperfusion in human liver. *Free Radical Research* **39**, 255–268.
- Chanson A, Sayd T, Rock E, Chambon C, Sante-Lhoutellier V, Potier de Courcey G & Brachet P (2005) Proteomic analysis reveals changes in the liver protein pattern of rats exposed to dietary folate deficiency. *Journal of Nutrition* **135**, 2524–2529.
- Dahlman L, Eriksson P, Kaaman M, Jiao H, Lindgren CM, Kere J & Arner P (2004) α 2-Heremans-Schmid glycoprotein gene polymorphisms are associated with adipocyte insulin action. *Diabetologia* **47**, 1974–1979.
- de Roos B, Duivenvoorden I, Rucklidge G, Reid M, Ross K, Lamers RJ, Voshol PJ, Havekes LM & Teusink B (2005a) Response of apolipoprotein E*3-Leiden transgenic mice to dietary fatty acids: combining liver proteomics with physiological data. *FASEB Journal* **19**, 813–815.
- de Roos B, Rucklidge G, Reid M, *et al.* (2005b) Divergent mechanisms of cis9, trans11- and trans10, cis12-conjugated linoleic acid affecting insulin resistance and inflammation in apolipoprotein E knockout mice. *FASEB Journal* **19**, 1746–1748.
- Diamandis EP (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *Journal of the National Cancer Institute* **96**, 353–356.
- Diks SH & Peppelenbosch MP (2004) Single cell proteomics for personalised medicine. *Trends in Molecular Medicine* **10**, 574–577.
- Feng Y, Zhu Y, Chen X, Sha J, Fan L & Chen Q (2005) Effects of diet-induced hypercholesterolemia on testosterone-regulated protein expression in mice liver. *Journal of Nanoscience and Nanotechnology* **5**, 1273–1276.
- Fliser D, Wittke S & Mischak H (2005) Capillary electrophoresis coupled to mass spectrometry for clinical diagnostic purposes. *Electrophoresis* **26**, 2708–2716.
- Galli RL, Shukitt-Hale B, Youdim KA & Joseph JA (2002) Fruit polyphenolics and brain aging: nutritional interventions targeting age-related neuronal and behavioral deficits. *Annals of the New York Academy of Science* **959**, 128–132.
- Go VL, Nguyen CT, Harris DM & Lee WN (2005) Nutrient-gene interaction: metabolic genotype-phenotype relationship. *Journal of Nutrition* **135**, 3016S–3020S.
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH & Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994–999.
- Gygi SP, Rist B, Griffin TJ, Eng J & Aebersold R (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *Journal of Proteome Research* **1**, 47–54.
- Haywood S, Simpson DM, Ross G & Beynon RJ (2005) The greater susceptibility of North Ronaldsay sheep compared with Cambridge sheep to copper-induced oxidative stress, mitochondrial damage and hepatic stellate cell activation. *Journal of Comparative Pathology* **133**, 114–127.

- Hooper L, Thompson RL, Harrison RA, *et al.* (2006) Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer: systematic review. *BMJ* **332**, 752–760.
- Issaq HJ, Veenstra TD, Conrads TP & Felschow D (2002) The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochemical and Biophysical Research Communications* **292**, 587–592.
- Kabuyama Y, Resing KA & Ahn NG (2004) Applying proteomics to signaling networks. *Current Opinion in Genetics and Development* **14**, 492–498.
- Kim H, Deshane J, Barnes S & Meleth S (2006) Proteomics analysis of the actions of grape seed extract in rat brain: technological and biological implications for the study of the actions of psychoactive compounds. *Life Sciences* **78**, 2060–2065.
- Kim J, Sun P, Lam YW, *et al.* (2005) Changes in serum proteomic patterns by presurgical α -tocopherol and L-selenomethionine supplementation in prostate cancer. *Cancer Epidemiology, Biomarkers and Prevention* **14**, 1697–1702.
- Kolkman A, Dirksem EH, Slijper M & Heck AJ (2005) Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labelling. *Molecular and Cellular Proteomics* **4**, 255–266.
- Marko-Varga GA, Nilsson J & Laurall T (2004) New directions of miniaturization within the biomarker research area. *Electrophoresis* **25**, 3479–3491.
- Martin SA, Vilhelmsson O, Medale F, Watt P, Kaushik S & Houlihan DF (2003) Proteomic sensitivity to dietary manipulations in rainbow trout. *Biochimica et Biophysica Acta* **1651**, 17–29.
- Mayr M, Chung YL, Mayr U, Yin X, Ly L, Troy H, Fredericks S, Hu Y, Griffiths JR & Xu Q (2005) Proteomic and metabolomic analyses of atherosclerotic vessels from apolipoprotein E-deficient mice reveal alterations in inflammation, oxidative stress, and energy metabolism. *Arteriosclerosis, Thrombosis and Vascular Biology* **25**, 2135–2142.
- Mitchell BL, Yasui Y, Lampe JW, Gafken PR & Lampe PD (2005) Evaluation of matrix-assisted laser desorption/ionization-time of flight mass spectrometry proteomic profiling: identification of α 2-HS glycoprotein B-chain as a biomarker of diet. *Proteomics* **5**, 2238–2246.
- Omenn GS, States DJ, Adamski M, *et al.* (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226–3245.
- Ong SE, Blagoev B, Kratchmarova K, Kristensen DB, Steen H, Pandey A & Ma M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cell Proteomics* **1**, 376–386.
- Papin J & Subramaniam S (2004) Bioinformatics and cellular signaling. *Current Opinion in Biotechnology* **15**, 78–81.
- Pappin DJ, Hojrup P & Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **3**, 327–332.
- Pavlickova P, Schneider E & Hug H (2004) Advances in recombinant antibody microarrays. *Clinica Chimica Acta* **343**, 17–35.
- Poon HF, Farr SA, Thongboonkeerd V, Lynn BC, Banks WA, Porley JE, Klein JB & Butterfield DA (2005) Proteomic analysis of specific brain proteins in aged SAMP8 mice treated with α -lipoic acid: implications for aging and age-related neurodegenerative disorders. *Neurochemistry International* **46**, 159–168.
- Prohaska JR (2000) Long-term functional consequences of malnutrition during brain development: copper. *Nutrition* **16**, 502–504.
- Rabilloud T (2003) Membrane proteins ride shotgun. *Nature Biotechnology* **21**, 508–510.
- Rai AJ, Stemmer PM, Zhang Z & Adams BL (2005) Analysis of Human Proteome Organization Plasma Proteome Project (HUPO PPP) reference specimens using surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry: multi-institution correlation of spectra and identification of biomarkers. *Proteomics* **5**, 3467–3474.
- Rayman MP (2005) Selenium in cancer prevention: a review of the evidence and mechanism of action. *Proceedings of the Nutrition Society* **64**, 527–542.
- Riggs L, Sioma C & Regnier FE (2001) Automated signature peptide approach for proteomics. *Journal of Chromatography* **924**, A359–A368.
- Righetti PG, Castagna A, Antonoli P & Boschetti E (2005) Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* **26**, 297–319.
- Rowell C, Carpenter DM & Lamartiniere CA (2005) Chemoprevention of breast cancer, proteomic discovery of genistein action in the rat mammary gland. *Journal of Nutrition* **135**, 2953S–2959S.
- Sagsted J, Bendixen E & Anderson HJ (2004) Identification of specific oxidatively modified proteins in chicken muscles using a combined immunologic and proteomic approach. *Journal of Agricultural and Food Chemistry* **16**, 3967–3974.
- Schmid GM, Converset V, Walter N, Sennitt MV, Leung KY, Byers H, Ward M, Hochstrasser DF, Cawthorne MA & Sanchez JC (2004) Effect of high-fat diet on the expression of proteins in muscle, adipose tissues, and liver of C57BL/6 mice. *Proteomics* **4**, 2270–2282.
- Schuschke DA (1997) Dietary copper in the physiology of the microcirculation. *Journal of Nutrition* **127**, 2274–2281.
- Stanley BA, Neverova I, Brown HA & Van Eyk JE (2003) Optimizing protein solubility for two-dimensional gel electrophoresis analysis of human myocardium. *Proteomics* **3**, 815.
- Thongboonkerd V, Chutipongtanate S, Kanlaya R, Songtawe N, Sinchaikul S, Parichatikanond P, Chen ST & Malasit P (2006) Proteomic identification of alterations in metabolic enzymes and signaling proteins in hypokalemic nephropathy. *Proteomics* **6**, 2273–2285.
- Tosco A, Siciliano RA, Cacace G, Mazzeo MF, Capone R, Malorni A, Leone A & Marzullo L (2005) Dietary effects of copper and iron deficiency on rat intestine: a differential display proteome analysis. *Journal of Proteome Research* **4**, 1781–1788.
- Traum AZ, Wells MP, Aivado M, Libermann TA, Ramoni MF & Schacter AD (2006) SELDI-TOF MS of quadruplicate urine and serum samples to evaluate changes related to storage conditions. *Proteomics* **6**, 1676–1680.
- Wang G, Wu WW, Zeng W, Chou CL & Shen RF (2006) Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: reproducibility, linearity, and application with complex proteomes. *Journal of Proteome Research* **5**, 1214–1223.
- Warnock DE, Fahy F & Taylor SW (2004) Identification of protein associations in organelles, using mass spectrometry-based proteomics. *Mass Spectrometry Reviews* **23**, 259–280.
- Weissingner EM, Nguyen-Khoa T, Fumeron C, Saltiel C, Walden M, Kaiser T, Mischak H, Druke TB, Lacour B & Massy ZA (2006) Effects of oral vitamin C supplementation in hemodialysis patients: a proteomic assessment. *Proteomics* **6**, 993–1000.
- Wolters DA, Washburn MP & Yates JR III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry* **73**, 5683–5690.
- Wu WW, Wang G, Baek SJ & Shen RF (2006) Comparative study of three proteomic quantitative methods, DIGE, cLCAT, and

- iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *Journal of Proteome Research* **5**, 651-658
- Xie H & Griffin TJ (2006) Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry. *Journal of Proteome Research* **5**, 1003-1009.
- Yates JR III (2004) Mass spectral analysis in proteomics. *Annual Review of Biophysics and Biomolecular Structure* **33**, 297-316.
- Zachara NE & Hart GW (2004) O-GlcNAc modification: a nutritional sensor that modulates proteasome function. *Trends in Cell Biology* **14**, 218-221.
- Zhang Y, Lu X, Shi Q, Xu X-Q, Leung HCE, Harris LN, Iglehart JD, Miron A, Liu JS & Wong WH (2006) Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *BMC Bioinformatics* **7**, 197-211.