Genomics and proteomics: importance for the future of nutrition research

H. Daniel*

Molecular Nutrition Unit, Department of Food and Nutrition, Technical University of Munich, Hochfeldweg 2, D-85350 Freising-Weihenstephan, Germany

A huge number of genes within the human genome code for proteins that mediate and/or control nutritional processes. Although a large body of information on the number of genes, on chromosomal localisation, gene structure and function has been gathered, we are far from understanding the orchestrated way of how they make metabolism. Nevertheless, based on the genetic information emerging on a daily basis, we are offered fantastic new tools that allow us new insights into the molecular basis of human metabolism under normal as well as pathophysiological conditions. Recent technological advancements have made it possible to analyse simultaneously large sets of mRNA and/or proteins expressed in a biological sample or to define genetic heterogeneity that may be important for the individual response of an organism to changes in its nutritional environment. Applications of the new techniques of genome and proteome analysis are central for the development of nutritional sciences in the next decade and its integration into the rapidly developing era of functional genomics.

Genome: Proteome: Nutrition research

Introduction

Human metabolism is mainly the expression of a steady state in the biosynthesis and degradation of proteins (turn over) that function either as enzymes, receptors, transporters, channels, hormones and other signalling molecules or that provide structural elements for cells, organs or the skeleton. Between the proteins there is a variable flow of metabolic intermediates that serve as building blocks for proteins, carbohydrates, lipids or hetero-oligomers and that provide the fuel for ATP synthesis. Whereas in the past we mainly looked at the phenotypical expression of metabolism by measuring concentrations and/or changes of metabolic intermediates, the new molecular tools allow us now to determine every step in the flow of the biological information from DNA to mRNA to proteins to function (Fig.1). Moreover, we now can assess the variability in the genetic make up (is 'code' the right word here? Is the genetic 'code' not universal?) and the expression level of RNA and proteins. Whereas *genomics* describes large scale DNA-sequencing that provides basic genetic information and insights into the heterogeneity in coding regions of genes (i.e. SNP: single nucleotide polymorphisms) as well as in control elements (i.e. promotors),

transcriptomics — also called expression profiling — assesses in a biological sample the mRNA-levels of a few or up to a several thousand genes simultaneously and this is mainly done by DNA-hybridisation arrays (Celis *et al.* 2000; Lockhart & Winzeler, 2000). *Proteomics* allows a huge number of proteins expressed in a cell or organ to be identified and changes in their expression pattern or level to be determined. Moreover, for individual proteins post-translational modifications that are crucial for functions or even amino acid substitutions (polymorphisms) can be detected (Dutt & Lee, 2000; Pandey & Mann, 2000).

Genomics employs either classical DNA-sequencer technology or more advanced techniques such as DNA-arrays or a 'gene balance' (see below). Transcriptome analysis utilises mainly fluorescence based detection systems to determine message (mRNA) expression levels in a biological sample by means of polymerase chain reaction (PCR) -techniques and Northern-blot analysis or by the sequence specific annealing of a immobilised capture oligonucleotide with the corresponding fragment obtained from a tissue sample on a DNA-microchip. Proteome analysis requires the isolation of proteins from a sample, their separation by two-dimensional polyacrylamide-gel

---

Abbreviations: IEF, isoelectric focusing; MALDI-TOFMS, matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry; PPAR, peroxisome proliferator activated receptor; SNP, single nucleotide polymorphism; 2D-PAGE, two dimensional polyacrylamide-gel electrophoresis.

Note: For the definition of the terms inulin and oligofructose please refer to the introductory paper (p. S139) and its footnote.

* **Corresponding author:** Prof. Dr H. Daniel, tel +49 8161 713 400, fax +49 8161 713 999, email daniel@weihenstephan.de

electrophoresis (2D-PAGE) and staining of the proteins in the gel. A computer-based comparison of gels (i.e. before and after treatment of cells or organisms) allows the pattern of protein expression to be determined. The identification of the proteins of interest showing increased or decreased expression levels requires the isolation of the protein from the gel, its digestion by trypsin (or another specific protease) with the generated peptide fragments then subjected to matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOFMS) analysis for determining the finger print of peptide masses that are characteristic for a given protein. Data base comparison with known DNA-sequences/amino acid sequences identifies the protein. Mass deviations of the measured peptide fragment mass from the expected amino acid sequence respectively its corresponding mass, may give indications for post-translational modifications such as phosphorylation, glycosylation or myristilation.

Taken together, applying these new techniques in nutritional sciences will not only extend our knowledge and understanding of basic mechanisms of gene expression control by nutritional factors, it also will provide new biomarkers that are required for studies in humans on the effects of life styles and nutrition as well as for assessing the safety and functionality of new foods or food ingredients.

Special technological applications in genomics

DNA-arrays

The general principle of any microarray technology is the hybridisation of oligonucleotides according to base pairing rules (i.e. A-T and G-C for DNA; A-U and G-C for RNA). The array provides a defined arrangement of samples using either microplates or standard blotting membranes or glass slides as support and can be made by hand or by using robotics to deposit the sample. Arrays may be manufactured as macroarrays or microarrays. Macroarrays usually have sample spot sizes of about 500 microns and can be easily imaged by existing gel and blot scanners. The sample spot sizes in microarrays are mostly less than 200 microns in diameter and these arrays contain thousands of spots and require specialised robotics for handling as well as advanced image analysis equipment. A single DNA chip can thus provide information on thousands of genes simultaneously and high throughput applications allow parallel gene expression and gene discovery studies for hundreds of samples.

The two major applications for the DNA-microarray technology include the identification of sequence variations (SNP, gene mutations) and the determination of the expression level (abundance) of a set of mRNA molecules. There are two variants of microarray technology based on the nature of an arrayed DNA sequence with known identity. In one assay system a probe cDNA between 500 and 5000 bases long is immobilised on a solid surface such as a glass slide using robotic based spotting and here the array is exposed to a set of target molecules either separately or in a mixture. In the other system an array of oligonucleotides probes (20 ∼ 25-mer oligos) is synthesised either directly on the chip or is immobilised on the chip after conventional synthesis. The array then is exposed to labelled DNA from the sample, hybridised, and the identity and/or abundance of the complementary sequences are determined.

As most biological molecules have naturally positive or negative charges, they move when exposed to an electric field to the area with the opposite charge. This principle is used in a special chip technology (Radtkey et al. 2000) that applies an electric current to individual test sites on an electronic microchip. This enables rapid movement and concentration of molecules and increases molecular binding on the microchip up to 1000 times compared to passive methods for hybridisation usually applied. Non-electronic (passive) technologies are not only slower but also prone to errors because of non-specific hybridisation of the DNA. Within the electric field hybridisation occurs faster and is extremely accurate by localising this DNA–DNA interaction. In preparation for testing of DNA samples, an electronic current is applied at selected locations (individual electrodes) on a blank microchip. Biotinylated samples are then drawn and bound to the charged test sites. This electronic addressing procedure can be done in multiple passes, until the chip is completely addressed. Once the chip is addressed, hybridisation experiments of many different kinds can be performed. Electronic stringency control allows the removal of unbound and nonspecifically bound DNA after hybridisation. Electronic multiplexing, where multiple test sites can be analysed simultaneously from a single sample, is a powerful application of this system. This is particularly important since multiple mutations, rather than single mutations, may be the cause of individual diseases.

DNA-analysis by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry

The recent developments in MALDI-TOFMS have made it possible to analyse biological molecules including oligonucleotides and DNA as an alternative to classical gel-based DNA sequencing (Yates, 2000; Ross et al. 2000). MALDI-TOFMS can detect DNA created by restriction endonuclease cleavage, primer extension products, mass-tagged oligonucleotides and genomic DNA. Since differences in the molecular weight (molecular mass) of the sample molecules are determined, one may call the mass spectrometer applied to genomics also a 'gene balance'. To measure the mass of molecules, the test material must be desolvated and ionised. The two mechanisms for ionisation of test molecules (DNA, peptides, proteins) are matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI). The MALDI technique is based on the incorporation of the test molecule into a matrix that is dried and the target molecules are then volatilised under vacuum by UV-laser irradiation. Only a very small fraction of the sample is ablated with each laser shot. Typically, a time-of-flight (TOF) mass analyser is coupled with MALDI and measures the elapsed time from acceleration of the ionised target molecules until it reaches the detector. MALDI-TOFMS allows very high-speed analyses with no requirement for labelling of DNA. However, one

problem for example in SNP (single nucleotide polymorphism)-analysis is the resolution. Mass differences between individual nucleotides, for example between A and T, are as low as 9 Da and make it difficult to unambiguously identify the nature of the nucleotide within a larger sequence. This problem may be overcome by a variety of new methods for the analysis of SNP by MALDI-TOFMS that have recently been described (Buetow *et al.* 2001; Sun *et al.* 2000). High throughput applications of MALDI-TOFMS with the most advanced mass spectrometers running fully automated allow up to 30 000 individual measurements per run and processing appears to be restricted only by handling of data gathered.

Special technological applications in proteomics

The term 'proteome' was introduced as the protein complement of the genome. Proteome analysis is based on the separation of proteins by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) which is still the protein separation method with the highest resolution available. Identifying a protein in the gel was however always an extremely difficult task. Mass spectrometry has been the key to solve this problem but getting information on the amino acid sequence based on the coding regions of identified genes has been equally important. The introduction of large-scale nucleotide sequencing of both expressed sequence tags (EST), as well as genomic DNA of different species provides all the information on the amino acid sequence of proteins and the corresponding peptides derived from a given protein by sequence specific cleavage. In addition, the development of mass spectrometers for ionisation and analysis of biological molecules based on their mass with very high accuracy is also important. Finally the development of computer algorithms able to match uninterpreted (or partially interpreted) mass spectra with translations of the nucleotide sequence from databases was central to the advancement of the field of proteomics.

The 2D-PAGE separates proteins according to charge (isoelectric point: pI) by isoelectric focusing (IEF) in the first dimension and according to size (molecular mass) by SDS-PAGE in the second dimension. It therefore has a unique capacity for the resolution of complex mixtures of proteins, permitting the simultaneous analysis of hundreds or even thousands of gene products (Gorg *et al.* 2000). However, not all proteins are resolved and separated equally well by 2D-PAGE. Very alkaline, hydrophobic and membrane proteins as well as high molecular weight proteins are still a problem. In some cases a prefractionation according to cellular compartment (membranes/ microsomes, cytosol, mitochondria) or by protein solubility by classical means may be necessary (Blackstock & Weir, 1999; Cordwell *et al.* 2000). In addition, proteins of low cellular abundance which may be particularly important in view of their cellular functions for example in signalling pathways are still very difficult to be resolved in the presence of large quantities of house-keeping proteins (Corthals *et al.* 2000). However, new concepts are constantly being developed that employ, for example tagging techniques (Adam *et al.* 2001) or the enrichment of the minor proteins prior to separation in 2D-gels.

Although 2D-PAGE may already appear as an 'ancient' technique, it is still not an easy attempt to obtain reproducible separation conditions and standardised gels. Similarly, the staining of the proteins in the gel is crucial for the quality of analysis afterwards. Both, Coomassie-blue as well as silver-staining can be performed with silver-staining generally giving a better signal to background ratio. Recently, sensitive fluorescent dyes have been developed that further improve the sensitivity (Patton, 2000). Assignments of protein spots in a gel or comparison of gels is performed by sophisticated computer software after scanning the gels with the programs also correcting the location of a protein spot in the gel based on the deviations in its position in the gel based on variances in separation quality.

It needs to be emphasised here that proteome analysis is straightforward if one assesses the effects of a treatment, for example in cultured cells or cell lines since here a homogenous population of cells is analysed. When tissue samples are utilised they contain different cell populations with different expression profiles contributing quite differently to the proteome pattern obtained. Variations in the biological sample (i.e. different percentages of individual cell types) consequently makes analysis a particularly difficult task and may require a separation of the different cell populations by means of cell specific surface markers and immunoaffinity techniques or by laser driven microdissection approaches (Banks *et al.* 1999; Simone *et al.* 2000).

Analysis of proteins of interest

Although the most common procedure for the identification of a protein spot in a gel is currently the peptide mapping or 'finger print' analysis, there are a number of other techniques and approaches that can also be applied (Gevaert & Vandekerckhove, 2000; Chalmers & Gaskell, 2000). For example, the proteins may be transferred from the gel onto a membrane to make them accessible for specific antibodies that then detect the proteins by the immunoreaction. Proteins can also, after excision from the gel, be submitted to classical Edman degradation with N-terminal sequencing to obtain an amino acid sequence that can be used to identify the protein by comparison with sequences deposited in data bases.

For obtaining sequence information the more rapid and advanced techniques employ the peptide mass analysis. Here the proteins are also excised from the gel and the gel is altered chemically to make the protein accessible to hydrolysis by a protease such as trypsin. Digestion requires a highly purified trypsin that should not contain any contaminating other enzymatic activity to have the specificity for cleavage of the peptide bonds containing cationic amino acid residues. Based on this site-specific hydrolysis a distinct and even characteristic pattern of peptide fragments of a given protein is serving as a peptide mass finger print.

The mixture of peptides isolated by digestion with the protease is usually submitted to MALDI-TOFMS since this technique with its recent advances allows peptides to be analysed directly by laser-driven transfer into the gas phase without or with only minor degradation during this
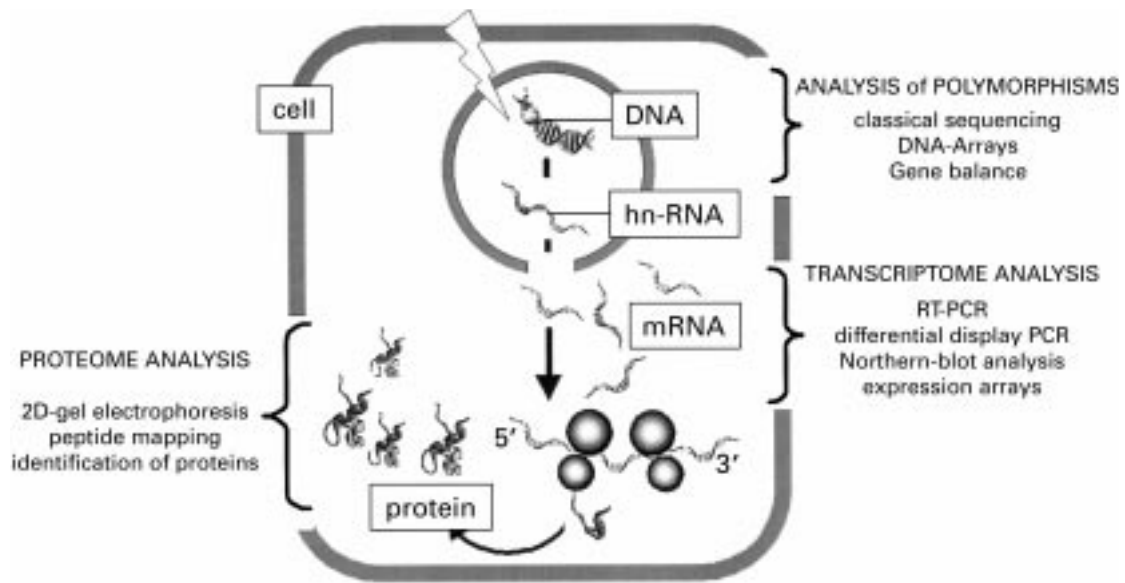
Fig. 1. The flow of biological information in a eukaryotic cell and the methods applied in genomics and proteomics to study the cellular response to changes in the nutritional environment.

process. Depending on the matrix used (mainly cinnamic acid or benzoic acid derivatives) a high resolution and accuracy in determination of the peptide masses is obtained. Although MALDI-TOFMS does not allow the amount of a protein or peptide to be determined, it provides a sensitivity of detection in the fmol-range and a precision

and reproducibility of masses in the ppm-range. The obtained mass spectrum is submitted to computer programs that by various algorithms interpret the pattern and predict the protein based on a comparison with masses predicted by 'virtual digestion' of identified open reading frames (coding amino acid sequences) in a given genome. Never-
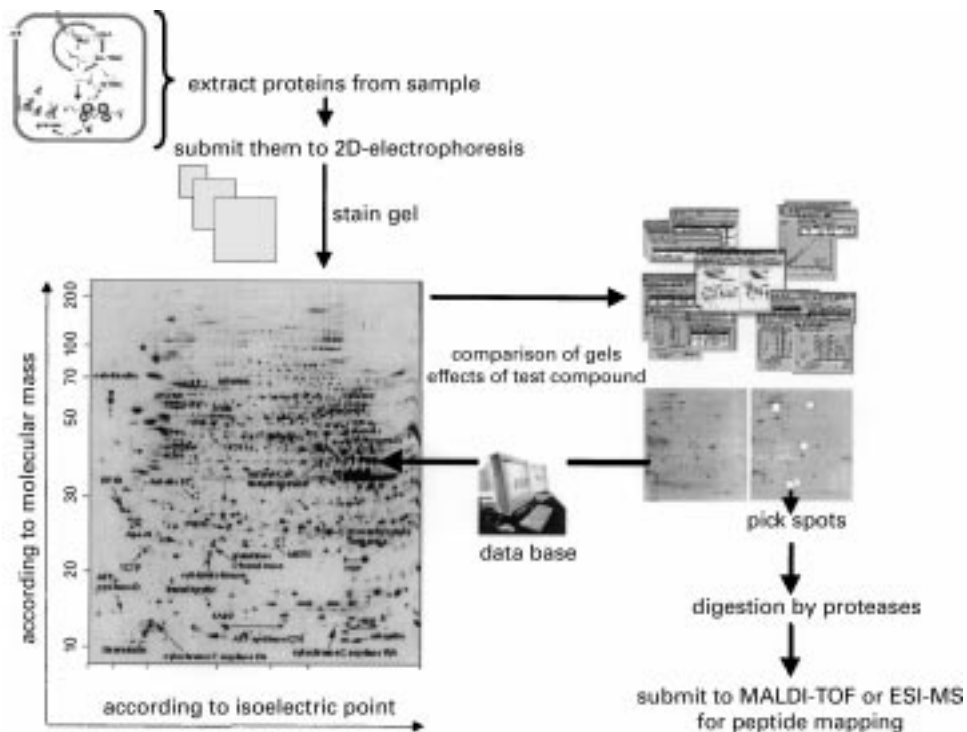

Fig. 2. Basic procedures for isolation of proteins from a biological sample by 2D-PAGE followed by computer based identification of protein spots in the gels and analysis by peptide mapping or other means.
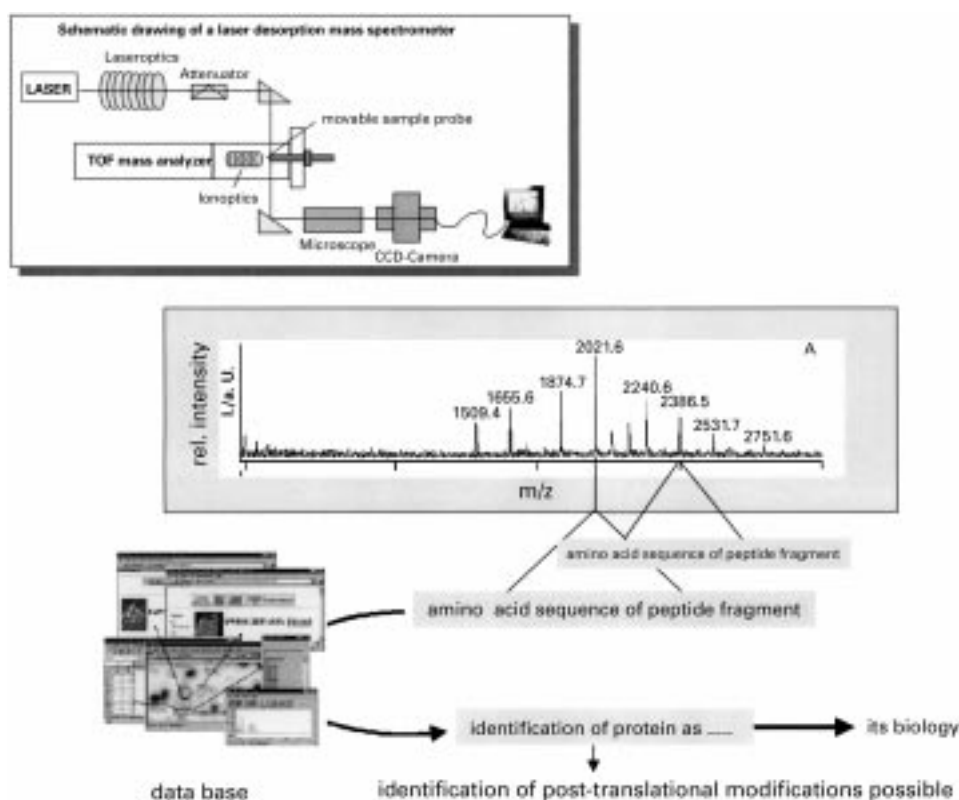
Fig. 3. Schematic of a MALDI-TOF mass analyser and the principle of determination and identification of a protein of interest based on a peptide mass spectrum

theless, predicted and measured mass spectra need interpretation in ensuring reproducibility and feasibility of the observation. In case of higher than predicted masses of a given peptide one could, based on the measured deviation, predict that post-translational modifications of the parent protein may have occurred such as addition of phosphate groups, hydroxylations at lysine or proline residues, glycosylations or addition of fatty acids. Other deviations of measured from predicted masses might be due to polymorphisms in coding sequences with subtle amino acid substitutions or even more pronounced with deletions or insertions. Such changes in the primary sequence can be resolved but strongly depend on the type of substitution (i.e. conservative exchange) and may require internal peptide Edman sequencing or more advanced mass analysis by electrospray ionisation (ESI) MS.

Applications of genomics and proteomics in nutritional sciences

Essentially every nutritional process relies on the interplay between a multitude of proteins encoded by a multitude of mRNA molecules expressed in a certain cell type or organ. Alterations of mRNA levels and in turn, of the corresponding proteins (although this not necessarily always changes in parallel) are critical parameters in controlling the flux of a nutrient or metabolite through a biochemical pathway. Nutrients and/or non-nutrient components of foods can affect gene expression either directly by interference with the control apparatus of the gene expression machinery or by virtue of metabolites or metabolic conditions induced (hormonal status, cellular redox status, etc.) that in turn directly or indirectly alter mRNA levels and/or the protein levels and thereby function. Application of genomics and proteomics technologies in studies on nutrition — either in cell cultures, animals or humans — has the great potential to identify specific markers (biomarkers) that respond to a certain nutrient, treatment or a diet. This may not mean that an individual gene by changes in its mRNA or protein level is identified as a 'true' marker, it may also be a pattern that changes in a characteristic way. So far biomarkers of cellular functions have been identified mainly by rational approaches based on the knowledge of metabolism. The new approach is essentially non-logical by its character as a screening process employing up to several thousands of potentially affected indicators of metabolic status.

Whereas the application of both technology platforms appears unlimited in terms of subjects of interest when utilising cells in culture or model organisms, there are clearly restrictions towards the use in studies in humans. This is mainly due to limited availability of cells for analysis. Although certain tissue cells (such as hair follicle cells, skin cells or even exfoliated intestinal cells) may be obtained in sufficient quantity, all different types of blood cells are generally an interesting source of material and may be used as 'reporter cells'. They do respond to dietary manoeuvres and more interestingly have different life times, different gene expression profiles and control

systems and reach and occupy different body compartments. As erythrocytes do not contain DNA and consequently respond at the protein level primarily only after new cells have been made one has a time-dependent integrated read out system with a lag time (although modifications and altered protein breaks may also occur in mature red cells). In particular, peripheral blood lymphocytes are already used for identification of potential mRNA biomarkers in human studies in response to environmental factors employing DNA-arrays (Glynne *et al.* 2000; Amundson *et al.* 2000; Marrack *et al.* 2000). For SNP-analysis different DNA sources can be used and large scale applications for identifications of relevant SNP are the basis of the rapidly growing field of molecular epidemiology in all areas of biomedical sciences (Perera & Weinstein, 2000; Tan *et al.* 1999; Haugen, 1999; Schork *et al.* 2000). Similarly, expression profiling, SNP-analysis and proteomics are already well established technology platforms in drug discovery studies and are central for pharmacogenomics, a phrase that covers genetic heterogeneity (SNP), selective expression of drug metabolising enzymes to phenotypical differences in the kinetics and dynamics of drug metabolism (Beeley *et al.* 2000; Rininger *et al.* 2000; Norton, 2001).

Examples of nutrigenomics

There are already a number of examples on applications of genomics and proteomics in studies on nutrition. The most impressive and probably the most expensive study demonstrating the power of 'nutritional genomics' has recently been published on the effects of ageing and caloric restriction in mice (Lee *et al.* 1999). The gene expression profiles of the ageing process were analysed in skeletal muscle by using high-density oligonucleotide arrays representing 6347 genes. One of the important findings is that only less then 1 % of those genes showed significant changes in expression level. Ageing resulted mainly in a differential expression pattern indicative of a marked stress response and lower expression of metabolic and biosynthetic genes. Most alterations were either completely or partially prevented by caloric restriction, the only intervention known to retard ageing in all mammals. Transcriptional patterns of animals on a low energy diet suggested that caloric restriction retards the ageing process by causing a metabolic shift toward increased protein turnover, increased gluconeogenesis and pentose-phosphate pathway and suppression of heat shock factors and inducible DNA-repair and detoxifying systems. A more recent study demonstrated similar characteristic changes in the transcriptom in mice brain by the ageing process (Lee *et al.* 2000).

An interesting example of proteome analysis applied in obesity research was recently published with the identification of drug targets of the peroxisome proliferator activated receptor (PPAR) transcription factors (Edvardsson *et al.* 1999). PPAR are ligand-activated receptors that regulate genes involved in lipid metabolism and energy homeostasis. The α-form of PPAR is preferentially expressed in the liver whereas the γ-form of PPAR is mainly found in adipose tissue. Activation of PPARα in rodents leads to peroxisome proliferation and increased β-oxidation of fatty acids. PPARγ-activation leads to adipocyte differentiation and improved insulin signalling of mature adipocytes. Consequently both PPAR-receptors are targets for treatment of obesity and lipidemia in man. ob/ob Mice, which have highly elevated levels of plasma triglycerides, glucose and insulin, serve frequently as a model of human obesity and insulin resistance. In search of drugs that interfere with PPARα, a specific agonist (WY14,643) was given to ob/ob mice and clinical improvements as well as proteome analysis of livers of mice were performed. After one week of treatment, plasma triglycerides, glucose and insulin levels in the mice were found to be significantly affected. The liver proteome pattern was determined by high-resolution two-dimensional gel electrophoresis which showed that WY14,643 produced up-regulation of at least sixteen protein spots. These proteins were identified by mass spectrometry and fourteen spots were found to represent proteins of the peroxisomal fatty acid metabolism. Those kind of studies are more or less already standard in commercial drug research as proteome analysis helps to find target proteins or metabolic pathways to be identified that show changes in protein expression related to wanted effects or unwanted side effects.

Nutrigenomics and nutriproteomics: which benefits for research on inulin and oligofructose?

As reviewed and discussed in this supplement, inulin and oligofructose have been shown to exert interesting nutritional effects. Genomics and proteomics will, in the future, help to understand the underlying mechanisms and, at the same time, help to characterise them better. This is certainly true for their effects on colonic epithelial cell functions including increased calcium absorption capacity and, eventually, better resistance to carcinogenic effects. The nutrient–gene interaction already identified as a possible causative event leading to hypotriglyceridemia will benefit from genomic as well as proteomic methodologies to investigate its relevance to human, especially in pathological situations like obesity, insulin resistance, etc. In the near future it is expected that genomics and/or proteomics will allow the identification and validation of new markers of cancer risk. The application of such markers will make possible new progress in the understanding and better characterising of the potential benefits of inulin and oligofructose consumption on colon cancer development, tumor growth and metastasis.

Summary

We are just entering a new era of postgenomic research and there is no doubt that nutritional science is going to be of central interest as nutrients and other food components, including inulin and oligofructose, are important factors in the gene-environment interactions. A wealth of the genetic information and novel techniques with high throughput capabilities provide fantastic new tools for nutrition research. Knowledge on the response of mammalian organisms to changes in their nutritional environment may be gathered at the mRNA and/or protein levels by

expression-arrays and proteome analysis (Trayhurn, 2000). As those tools can generate overwhelming data sets there is clearly an emerging need for bioinformatics in nutritional sciences but also for Nutritionists with a good background in cell biology and biochemistry of metabolism.

References

Adam GC, Cravatt BF & Sorensen EJ (2001) Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chemistry and Biology* **8**, 81–95.

Amundson SA, Do KT, Shahab S, Bittner M, Meltzer P, Trent J & Fornace AJ Jr (2000) Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiation Research* **154**, 342–346.

Banks RE, Dunn MJ, Forbes MA, Stanley A, Pappin D, Naven T, Gough M, Harnden P & Selby PJ (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis-preliminary findings. *Electrophoresis* **20**, 689–700.

Beeley LJ, Duckworth DM & Southan C (2000) The impact of genomics on drug discovery. *Progress in Medicinal Chemistry* **37**, 1–43.

Blackstock WP & Weir MP (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotechnology* **17**, 121–127.

Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR & Braun A (2001) High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences USA* **98**, 581–584.

Celis JE, Kruhoffer M, Gromova I, Frederiksen C, Ostergaard M, Thykjaer T, Gromov P, Yu J, Palsdottir H, Magnusson N & Orntoft TF (2000) Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Letters* **480**, 2–16.

Chalmers MJ & Gaskell SJ (2000) Advances in mass spectrometry for proteome analysis. *Current Opinion in Biotechnology* **11**, 384–390.

Cordwell SJ, Nouwens AS, Verrills NM, Basseal DJ & Walsh BJ (2000) Subproteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. *Electrophoresis* **21**, 1094–1103.

Corthals GL, Wasinger VC, Hochstrasser DF & Sanchez JC (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21**, 1104–1115.

Dutt MJ & Lee KH (2000) Proteomic analysis. *Current Opinion in Biotechnology* **11**, 176–179.

Edvardsson U, Alexandersson M, Brockenhuus von Lowenhielm H, Nystrom AC, Ljung B, Nilsson F & Dahllof B (1999) A proteome analysis of livers from obese (ob/ob) mice treated with the peroxisome proliferator WY14,643. *Electrophoresis* **20**, 935–942.

Gevaert K & Vandekerckhove J (2000) Protein identification methods in proteomics. *Electrophoresis* **21**, 1145–1154.

Glynne R, Ghandour G, Rayner J, Mack DH & Goodnow CC (2000) B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays. *Immunolical Reviews* **176**, 216–246.

Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R & Weiss W (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037–1053.

Haugen A (1999) Progress and potential of genetic susceptibility to environmental toxicants. *Scandinavian Journal of Work Environment and Health* **25**, 537–540.

Lee CK, Klopp RG, Weindruch R & Prolla TA (1999) Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**, 1390–1393.

Lee CK, Weindruch R & Prolla TA (2000) Gene-expression profile of the ageing brain in mice. *Nature Genetics* **25**, 294–297.

Lockhart DJ & Winzeler EA (2000) Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836.

Marrack P, Mitchell T, Hildeman D, Kedl R, Teague TK, Bender J, Rees W, Schaefer BC & Kappler J (2000) Genomic-scale analysis of gene expression in resting and activated T cells. *Current Opinion in Immunology* **12**, 206–209.

Norton RM (2001) Clinical pharmacogenomics: applications in pharmaceutical R&D. *Drug Discovery Today* **6**, 180–185.

Pandey A & Mann M (2000) Proteomics to study genes and genomes. *Nature* **405**, 837–846.

Patton WF (2000) Making blind robots see: the synergy between fluorescent dyes and imaging devices in automated proteomics. *Biotechniques* **28**, 944–948.

Perera FP & Weinstein IB (2000) Molecular epidemiology: recent advances and future directions. *Carcinogenesis* **21**, 517–524.

Radtkey R, Feng L, Muralhidar M, Duhon M, Canter D, DiPierro D, Fallon S, Tu E, McElfresh K, Nerenberg M & Sosnowski R (2000) Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. *Nucleic Acids Research* **28**, E17–E17.

Rininger JA, DiPippo VA & Gould-Rothberg BE (2000) Differential gene expression technologies for identifying surrogate markers of drug efficacy and toxicity. *Drug Discovery Today* **5**, 560–568.

Ross P, Hall L & Haff LA (2000) Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. *Biotechniques* **29**, 620–626.

Schork NJ, Fallin D & Lanchbury JS (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. *Clinical Genetics* **58**, 250–264.

Simone NL, Paweletz CP, Charboneau L, Petricoin EF & Liotta LA (2000) Laser capture microdissection: beyond functional genomics to proteomics. *Molecular Diagnosis* **5**, 301–307.

Sun X, Ding H, Hung K & Guo B (2000) A new MALDI-TOF based mini-sequencing assay for genotyping of SNPS. *Nucleic Acids Research* **28**, E68–E68.

Tan KT, Dempsey A & Liew CC (1999) Cardiac genes and gene databases for cardiovascular disease genetics. *Current Hypertension Report* **1**, 51–58.

Trayhurn P (2000) Proteomics and nutrition: a science for the first decade of the new millennium. *British Journal of Nutrition* **83**, 1–2.

Yates JR III (2000) Mass spectrometry. From genomics to proteomics. *Trends in Genetics* **16**, 5–8.