Minireview

# Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics

Julio E. Celis[a],*, Mogens Kruhøffer[b], Irina Gromova[a], Casper Frederiksen[b],
Morten Østergaard[a], Thomas Thykjaer[b], Pavel Gromov[a], Jinsheng Yu[b], Hildur Pálsdóttir[a],
Nils Magnusson[a], Torben F. Ørntoft[b,1]

[a]*Department of Medical Biochemistry and Danish Centre for Human Genome Research, University of Aarhus, Ole Worms Alle, Bldg. 170,
DK-8000 Aarhus C, Denmark*
[b]*Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark*

**Abstract** Novel and powerful technologies such as DNA microarrays and proteomics have made possible the analysis of the expression levels of multiple genes simultaneously both in health and disease. In combination, these technologies promise to revolutionize biology, in particular in the area of molecular medicine as they are expected to reveal gene regulation events involved in disease progression as well as to pinpoint potential targets for drug discovery and diagnostics. Here, we review the current status of these technologies and highlight some studies in which they have been applied in concert to the analysis of biopsy specimens. © 2000 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

*Key words:* Gene expression profile; Microarray; Proteome; Two-dimensional gel electrophoresis; Immunohistochemistry; Bladder cancer

## 1. Introduction

It is now a little more than 10 years since the Human Genome Project was launched and during this relatively short period of time there have been remarkable advances in the construction of physical and genetic maps as well as in the identification of genes associated with human diseases [1,2]. Soon, the total sequence of the human genome will be deciphered and, hopefully, made available to researchers worldwide for the benefit of mankind.

Undoubtedly, the Human Genome Project has paved the way to the revolution in the life sciences that we are experiencing today. Gradually, however, its focus is starting to shift towards functional genomics, an area of the post-genomic era that deals with the functional analysis of genes and their products (see the article by Goffeau in this issue). Techniques of functional genomics include methods for gene expression profiling at the transcript (DNA microarrays [3–5]; see also article by Brazma and Vilo in this issue); differential display [6]; serial analysis of gene expression [7–9], and protein levels (proteomics) ([10–12]; see also article by Andersen and Mann in this issue), as well as transgenics [13], phage display [14], procedures for studying protein–protein interactions ([15,16]; see also article by Legrain and Selig in this issue) and bioinformatics [17].

Among the techniques of functional genomics, both DNA microarrays ([3–5] and references therein) and proteomics [10–12] hold great promise for the study of complex biological systems with applications in molecular medicine. These novel and powerful gene expression profiling techniques permit the analysis of the expression levels of thousands of genes simultaneously both in health and disease. These technologies are complementary, allow high-throughput, and in combination are expected to generate a vast amount of gene and protein expression data that may lead to a better understanding of the regulatory events involved in normal and disease processes. In addition, these technologies offer a systematic approach for searching for effective targets for drug discovery and diagnostics.

Here, we review the current state of DNA microarrays and proteomics and highlight some studies in which they have been applied in concert to the analysis of biopsy specimens.

## 2. DNA microarrays

The amount of information that is now becoming available to researchers in the life sciences is exploding, and even though the data can be stored in conventional media, new methods are being required to analyze large sets of genes in a high-throughput fashion. For this purpose the DNA array technology was developed. The method makes it possible to survey thousands of genes in parallel, and has several areas of application. One is expression monitoring [18], in which the transcript levels of genes are measured in different physiological conditions both in cultured cells and tissues, to search for regulatory expression patterns. Understanding patterns of expressed genes is expected to improve our knowledge of highly complex networks that cross communicate in hitherto unknown ways both in health and disease. Another area of application is polymorphism analysis [19]. In this case, polymorphic regions of the genome are scanned to search for link-

*Corresponding author. Fax: (45)-8613-1160.
E-mail: jec@biokemi.au.dk

[1] Also corresponding author. E-mail: orntoft@kba.sks.au.dk

age to diseases, and to reveal disease susceptibility genes and/ or inherited disease genes. A similar approach has been used to analyze polymorphic regions of known genes, in particular to determine whether polymorphisms are associated with an altered function of the gene product, a fact that may increase the susceptibility to disease. Finally, various attempts have been made to utilize DNA arrays for sequencing [20].

Below we review the technology currently in use for microarray-generated gene expression pattern discovery as well as some applications. Table 1 provides a set of World Wide Web sites that contain useful additional information (see also the article by Brazma and Vilo in this issue).

## 2.1. Microarray technology

Microarrays are usually made by deposition of DNA spots on a solid support like a coated glass surface, that differs in several ways from conventional filter-based supports such as charged nylon and nitrocellulose. The flatness of the glass surface makes it possible (i) to array molecules in a parallel fashion, (ii) to miniaturize the procedure and (iii) to use fluorescent dyes for detection. There is no diffusion of the applied material into the support, thus allowing focusing for laser scanning microscopy.

Two main procedures have been used to produce DNA chips: photolithography as developed and marketed primarily by Affymetrix Inc. (Santa Clara, CA, USA) [20,21] and mechanical gridding [22]. Photolithography is well known in the computer chip industry and utilizes an ultraviolet light source that passes through a mask that directs in a step-wise manner where a photochemical reaction (oligonucleotide synthesis) takes place on a siliconized glass surface. The mask can be produced with openings as small as a few micrometers allowing a density of several hundred thousand probes per square centimeter of glass. There is, however, an inherent length restriction with this in situ synthesis technology limiting the probes to about 25 nucleotides in length. High-density arrays, on the other hand, allow the use of multiple probes per gene [20].

Mechanical gridding methods are based on ink-jet or physical deposition of the material using pins manufactured with very high precision. There is direct surface contact and the transport of small amounts of liquid makes these systems susceptible to evaporation and contamination with dust particles. The gridding instruments use an XYZ motion control based on step engines that can be controlled with very high precision. The DNA containing material can be spotted from 96 or 384 well plates to glass in predefined patterns.

The arrayed probes can be oligos (photolithography and gridding) or cDNAs (gridding). The hybridization reaction conditions are quite different in the two cases, and special sample preparations are needed to optimally utilize these probes. With shorter probes, i.e. of 20–50 nucleotides, the sample is fragmented to avoid tertiary structures and to achieve optimal hybridization [18,20]. Polymerase chain reaction (PCR) amplified probes of 300–2000 nucleotides usually do not require fragmentation of the sample [23].

The type of glass used as support, the coating substance, the coupling technique, the labelling system, and the fluorescent labels used for detection are all variables that must be optimized. A number of coating substances are commercially available to immobilize DNA to the surface. Two types of slides are available that use a coating procedure designed for printing of amine modified DNA: these include (i) silylated slides, which contain reactive aldehyde groups that react with amino-groups via a Schiff base formation as the printed DNA dries on the surface of the slide (available from Telechem Inc., CA, USA; Cell Associates Inc., TX, USA; for coupling chemistry see http//www.arrayit.com/microarraycoupling/) and (ii) activated slides prepared by the covalent attachment of a hydrophilic, polymeric amine reactive coating to silane base-coated slides as described by Beier and Hoheisel [24] (available from Surmodics Inc., MS, USA). Amine-modified DNA attaches covalently to the activated polymeric surface.

Three other types of slides are available that are based on more conventional immobilization technology routinely used in membrane immobilization: these include (i) silanized slides which carry covalently attached primary amines on the surface that can form ionic bonds with the phosphate backbone at neutral pH. In addition, the radical-based coupling between thymidine residues on the DNA and carbons on the alkyl amine of the substrate can be induced with UV or heat (provided by companies like Telechem Inc., CA, USA; Sigma Aldrich, Inc., MO, USA; Corning, Inc., NY, USA), (ii) nitrocellulose-based polymer-coated slides that possess the binding and immobilization properties of nitrocellulose that binds DNA in a non-covalent but irreversible manner (Schleicher and Schuell), and (iii) poly-lysine-coated slides which require UV crosslinking of the DNA (available from Sigma Aldrich Inc., MO, USA; see also http//cmgm.stanford.edu/pbrown/protocols/1_slides.html).

## 2.2. Quantitating the signals from arrays

A linear response that covers two or three orders of magnitude is often needed to detect low and high copy number transcripts on the same array. In cases where this is not possible it may be necessary to scan the chip at different wavelengths, or to amplify the signal with an immune sandwich on top of the bound sample [25]. In the latter case, the first scanning is carried out after hybridization of the labelled sample, and the second after reaction with the labelled antibodies.

It is necessary to document the linearity and reproducibility in each step of the procedure, and sometimes even from probe to probe to obtain reliable data. Often, a standard sample is used to compare with the experimental sample and this may compensate for differences in hybridization from probe to probe.

## 2.3. Standardization

Comparison of data obtained from independent arrays and from different laboratories requires standardization. Both the Affymetrix chips and the custom made cDNA chips use different methods for standardization. The Affymetrix chips have approximately 20 probes per gene and standardization is either based on the expression level of selected genes, like actin and GAPDH, or on a setting of the global chip intensity to approximately 150 units per gene on the chip. In this way, chip data from different experiments can be compared to each other. In our hands, the data obtained with the two standardization methods differ only by approximately 10% (unpublished observations).

The custom-made cDNA or oligo arrays also require standardization, but this is a complex problem. In general, the standard used often reflects the purpose for which the array

was produced. For example, for expression monitoring of breast cancer cells, a mixture of breast cancer cell lines may provide a good standard [26]. Today, however, there is no golden standard that can be used for all purposes and as a result, it is difficult to compare data from different laboratories and often it is necessary to use other technologies such as Northern hybridization, real-time PCR or immunostaining to validate the signals. A minimum requirement is that laboratories that produce arrays themselves should be able to reproduce data from one chip generation to the next based on the analysis of well-defined controls covering different genes and expression levels. In addition, it is common to use spiking of samples with bacterial genes that hybridize to probes spotted for control purpose on the arrays.

### 2.4. Samples for expression monitoring

The analysis of relatively homogeneous cell populations (cloned cell lines, yeast, etc.) has proven much simpler than the analysis of tissue biopsies as the latter often contain many cell types (epithelial, endothelial, inflammatory, nerve, muscle, and connective tissue cells) that are present in variable amounts. Standardization may require microdissection of the tissue to isolate specific cell types [27,28], although the number of cells needed for the assay is well above a million. Sampling of specific cell types using laser capture microdissection (LCM) [29] can be a time-consuming task, and given that mRNA is prone to degradation the processing time must be kept to a minimum. If only a small amount of material is available, then a reverse transcription-PCR step is necessary for amplification, but this adds an additional complication due to the lack of linear amplification of all transcripts. In one of our laboratories we have used preparations of single cell types from tumor biopsies to standardize pooling of samples for generating profiles of gene expression at different stages of tumor development (manuscript in preparation).

### 2.5. Bioinformatic analysis of expression data

*2.5.1. Hierarchical cluster algorithms.* Hierarchical clustering algorithms can be divided into two types: agglomerative and divisive (Fig. 1) [30]. The agglomerative method is a bottom-up approach, where the algorithm starts with $n$ separate clusters (for example 4000 genes, where $n = 4000$) and successively combines clusters until only one is left. The divisive method, on the other hand, is a top-down approach starting with one cluster and successively splitting clusters to produce others. The algorithm used to form the clusters must also be defined: two widely used and simple algorithms are the single linkage and average linkage methods, respectively. Single linkage, also called nearest neighbor, defines the distance between two clusters as the minimum distance over all pairs of clusters. Average linkage takes in consideration the average distance over all [30].

A distance matrix must be calculated before clustering is performed, and it is the distances between the obtained gene expression profiles that are used to form the actual clusters [30]. Observations with small distances are grouped together as described above. The two most commonly used distance measurements are the Euclidean distance and the Pearson correlation coefficient.

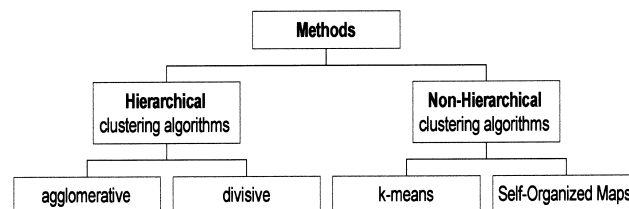Euclidean distance : $ED = \sqrt{(X_{ik} - X_{jk})^2}$



Fig. 1. Two main types of clustering algorithms, the hierarchical and the non-hierarchical algorithms.

Pearson correlation coefficient:

$$\text{Pearson} = \sqrt{\frac{(X_{ik} - X_{jk})^2}{V_k}}$$

$X_{ik}$ is the measurement for the $k$th variable on sampling unit $i$. $V_k$ is the variance of the $k$th variable [30,31].

*2.5.2. Non-hierarchical cluster algorithms.* In non-hierarchical cluster analysis it is assumed that the data can be divided into a certain number of clusters and that they are well separated. The advantage of this approach is that large data sets can be clustered much faster than by using hierarchical clustering because a lower number of clusters is assigned. The most common method for non-hierarchical cluster analysis is $k$-means. However, a method termed self-organized maps (SOM) has recently been applied to expression data generated from DNA chip arrays [32,33]. The $k$-means method [34] identifies $k$ points that function as cluster centers. Each data point is then assigned to one of these centers in a way that minimizes the sum of the distances between all points and their centers. Thus, it is the distribution of points that decides the value of the means. One drawback of this method is that a specific number of clusters is assigned, as the number of clusters is usually unknown in large data sets. SOM is similar to the $k$-means approach, but it has a geometrical configuration and the number of nodes predefines this configuration. Initially, data points are mapped onto the geometrical configuration. When clustering the data with SOM the position of a node migrates to fit the data points during successive iterations [32,33].

*2.5.3. Supervised classification.* Common to the clustering methods in which array data are used is that they are unsupervised, i.e. no predefined references are known. An alternative option is to construct a supervised classification method that requires at least two references. For cancer classification, for example, the references could be the gene expression profiles from normal and invasive tumor tissue. In this particular case, a vector representing gene expression over $n$ genes can be used to describe each tissue [35,36]. In a recent article by Golub and co-workers [36], the authors analyzed 6817 genes using 38 bone marrow samples. Based on these 38 samples they found that a vector based on 10 and 200 genes was sufficient to distinguish between acute myeloid leukemia and acute lymphoblastic leukemia. Thus, the authors were able to construct a cancer classifier based on a low number of genes. Using a similar vector-based classifier approach, Brown and coworkers [37] analyzed 2467 yeast genes in 79 different experiments and were able to classify genes into functional categories based on the expression data from DNA chip arrays.

## 2.6. Applications of arrays for expression monitoring

One of the main areas for array application is in the simultaneous monitoring of thousands of transcripts in different biological settings. The approach is being used to identify new networks and to understand patterns of expressed genes. A number of articles have been published using the array technology aiming at identifying disease-associated alterations in humans. For this purpose clinical samples, human cell lines, and in a few cases animal models of human disease have been used.

*2.6.1. Gene expression profiling of tissue biopsies, cell lines and animal models of disease. 2.6.1.1. Tissue biopsies.* Gene expression studies on clinical samples have been performed in breast and colon cancer [26,38], as well as in atherosclerosis [39,40]. Genes of presumably known functions have been identified and linked to the diseases; most of these data are now available in the Internet (Table 1). In a study of breast cancer that used clinical specimens and cell lines, with an array containing approximately 5000 genes, Perou and colleagues [26] identified a proliferation related gene cluster in the cell line that was upregulated in the more aggressive clinical breast tumor specimens [26]. In a similar study of colon cancer based on the Affymetrix arrays, 48 EST's homologous to ribosomal proteins were found to be unregulated in the tumor tissue [38]. In this study, a muscle index was used for correcting for the stromal components as this showed a high index in the normal biopsies. Both the breast and the colon cancer studies lacked a correlation between gene expression levels in the cell lines and in the clinical specimens.

Human atherosclerosis lesions from arteria carotis sampled from patients undergoing surgery have also been analyzed using the Affymetrix expression arrays [40]. One important finding was a five-fold upregulation of the early growth response gene Egr-1, a DNA binding protein that influences the transcription of genes encoding growth factors, cytokines, adhesion molecules, and proteins related to coagulation. This finding was corroborated by immunohistochemistry and animal experiments, and identified Egr-1 as a possible target for therapeutic intervention.

*2.6.1.2. Cell lines.* Several studies based on expression monitoring have been performed in human cell lines in areas as diverse as cancer [41], ophthalmology [42], and the central nervous system [43]. Arrays have also been used to study the effect of cytokines (interferons) [44], cytomegalovirus infection [45], and oncogene transfection [46] on the overall patterns of gene expression. In one study, human foreskin fibroblasts were infected with human cytomegalovirus and the expression of approximately 6000 genes was monitored for up to 24 h [45]. A total of 258 genes was found to be upregulated more than four-fold. These included HLA-E (upregulated six-fold),

a protein that protects against cytotoxic T-lymphocytes; RO/SSA (52 kDa protein mRNA; upregulated 12-fold), a commonly targeted autoantigen, as well as several components of the pathway that produce prostaglandin E2 from arachidonic acid. In another study, it was shown that treatment of the human fibrosarcoma cell line HT 1080 with IFNs α, β and γ resulted in the upregulation of novel genes implicated in apoptosis (RAP46, Bag-1, scramblase), while genes like IGF-2 and ZnT-3 were strongly downregulated [44].

The effect on gene expression of the fusion oncogene PAX3-FKHR transfected into NIH 3T3 cells has also been studied using microarrays [46]. It was shown that the fusion gene, but not the wild type control was able to activate a myogenic transcription program that included induction of a number of transcription factors such as MyoD, myogenin, Six1, and Slug.

*2.6.1.3. Animal model systems.* A common approach to the study of human diseases is to use animal model systems. This has been done for a range of diseases including encephalomyelopathy, lymphoma, renal tubuli [23], and lung fibrosis [47]. In the latter case, Affymetrix chips were probed with pooled samples obtained from groups of six animals to reduce variations and cost. Transcripts that were significantly altered in lung fibrosis [47] included extracellular matrix and inflammatory response genes. Analysis of clusters containing these genes in a time course experiment with the SPOTFIRE PRO 3.0 program revealed different temporal patterns of expression that further subdivided these genes.

*2.6.2. Toxicology and drug testing.* From a toxicological point of view, there are great expectations for expression monitoring as the effect of drugs, both expected and unforeseen side effects, can be monitored in animals and eventually in humans [48,49]. One problem that has arisen from these studies, even in quite simple model systems, is the often-unexplained changes in transcript levels observed. These changes are quite reproducible, suggesting a much more complex relationship among gene products than previously thought [49].

The combination of gene expression monitoring and testing of drugs on cell lines and in animal models holds great promise. A recent publication showed that the variation in the liver expression of genes encoding xenobiotic metabolizing enzymes, glutathione regulators, DNA repair enzymes, heat shock proteins and housekeeping genes is larger among individual animals than that introduced by the array assay itself [49]. These studies revealed the upregulation of cytochromes even at low doses of β-naphthoflavone treatment and showed a good correlation between the array and Northern hybridization data. A recent screening using cDNA microarrays of 60 human cell lines used by the National Cancer Institute for

Table 1
Useful links on the World Wide Web for array data and software for data analysis

| Ref. | Link | Available |
| --- | --- | --- |
| [1] | http://llmpp. nih.gov/lymphoma/ | data and software |
| [3] | http://www.cse.ucsc.edu/research/compbio/genex/genex.html | data and software |
| [6,12,17,18] | http://www.genome.wi.mit.edu/MPR/data_sets.html | data and software (GeneCluster) |
| [11] | http://www.rii.com/tech/pubs/nm41293.htm | data |
| [13] | http://arep.med.harvard.edu/network_discovery/ | data (results) |
| [14] | http://quantgen.stanford.edu/ | data (results) |

drug discovery, has formed the basis for establishing a database that can be linked to gene expression data and molecular pharmacology [41,50].

*2.6.3. Molecular classification of diseases.* Classification of diseases can generally be achieved by class discovery and class prediction [36]. Class discovery refers to the identification, based on gene expression, of previously unrecognized subtypes of the disease. If clinical follow-up material is available these classes can be related to signs and symptoms, disease course, treatment outcome, and mortality. Class prediction refers to the ability to assign a particular patient to an already defined class, based on molecular examination of the diseased tissue or other specimens.

Two recent examples have demonstrated the power of expression arrays to classify hematological malignancies. In one of these studies, leukemias were classified as AML and ALL, respectively, based on 50 genes selected from an array carrying 6817 genes. In this study, 36 out of 38 patients were correctly classified, and two were uncertain. The 50 genes used in the array included some that were known to differ between AML and ALL, as well as new markers [36].

Another study classified B-cell lymphomas into two molecularly distinct classes that reflected different stages of B-cell differentiation, the germinal center-like and activated B-cell like lymphoma groups, respectively. For this study the lymphochip microarray from Research Genetics, which holds 17856 genes, was used. Follow-up for 12 years showed a significant difference in survival among these groups [51].

It is envisaged that global surveys of gene expression will identify marker genes that may be used to group patients into molecularly relevant categories; these markers are expected to greatly improve the precision and power of clinical trials.

### 2.7. Conclusions

There is no doubt that the massive parallel gene expression information generated by microarrays will have a major impact in the discovery and understanding of patterns of expressed genes. In addition, the technology is expected to generate novel and effective targets for drug discovery and provide, in combination with proteomics, valuable tools for the entire process of drug development and evaluation.

One of the main challenges we foresee in the future will be to solve the problems posed by the analysis, interpretation and access to the large amount of information that will be generated. Large studies of EST's have not been published yet as data analysis requires the development of new bioinformatic tools that can deal with the huge amount of information that is being created. These studies are expected to identify new genes of importance to specific biological processes and reveal new regulatory pathways through the analysis of the expression levels of individual EST's in large numbers of samples.

In the future, coupling of expression monitoring to transgenic animal models may prove to be quite rewarding as the global effect of a gene knock-out or knock-in can be monitored with both microarrays and proteomics tools [10].

## 3. Proteomics

A complementary technology to DNA microarrays for monitoring gene expression is provided by proteomics, a term generally used to encapsulate all of the technology currently available to analyze global patterns of gene expression at the protein level. Proteins are frequently the functional molecules and, therefore, the most likely to reflect differences in gene expression. Genes may be present, they may be mutated, but they are not necessarily transcribed. Some messengers are transcribed but not translated, and the number of mRNA copies does not necessarily reflect the number of functional protein molecules [52]. In addition, proteomics addresses problems that cannot be approached by DNA analysis, namely, relative abundance of the protein product, post-translational modification, subcellular localization, turnover, interaction with other proteins as well as functional aspects.

The proteome has been defined by Wilkins and colleagues as the complete set of **prote**ins encoded by the gen**ome** [53], and recently, the term has been broadened to include the set of proteins expressed both in space and time. There are two main approaches to proteomics: one is the expression model in which all proteins are analyzed, and the other is the cell map model in which only a selected set of proteins, like complexes and organelles, are studied [54].

The proteomic technology is complex, and comprises a plethora of state of the art techniques to resolve (high resolution two-dimensional gels), quantitate (phosphorimager, special scanners), identify and characterize proteins (microsequencing, mass spectrometry), as well as to store (two-dimensional polyacrylamide gel electrophoresis (2D PAGE) databases; http://biobase/dk/cgi-bin/celis; http://expasy.hcuge.ch/sprot/sprot-top.html) communicate and interlink protein and DNA sequence and mapping information (bioinformatics) [10,12,55–58]

The usefulness of the 2D PAGE technique for large-scale proteomic projects depends very much on the number of proteins that can be resolved in a complex protein mixture, for example a human cell. Proteome profile data from a few laboratories, including one of our own, have indicated that only a fraction of the human genes are switched-on in a given cell type and extensive analysis of whole cell extracts, organelles as well as partially purified subcellular fractions, suggest that individual cells may not express more than 6000 primary translation products [59]. To this number one has to add the post-translational processing and chemical modifications (phosphorylation, glycosylation, demethylation, acetylation, myristoylation, palmitoylation, sulfation, ubiquination, etc.), the latter being rather common and extensive in many proteins [60,61]. Thus, as we stand today the 2D PAGE technology is not able to resolve and depict in a single gel all of the proteins thought to be present in a mammalian cell.

Currently, there is a great deal of interest in proteomics as applications of this technology are expected to reveal gene regulation events involved in disease progression as well as potential targets for drug discovery and diagnostics. Moreover, the technology is bound to have a great impact in agriculture, toxicology and the industry in general.

### 3.1. Protein separation: the 2D PAGE technology

For the past 24 years, high resolution 2D PAGE has been the technique of choice for analyzing the protein composition of cells, tissues and fluids, as well as for studying changes in global patterns of gene expression elicited by a wide array of effectors [12,62–64]. The technique, which was originally described by O'Farrell [65,66] and Klose [67], separates proteins both in terms of their isoelectric point (p*I*) and molecular
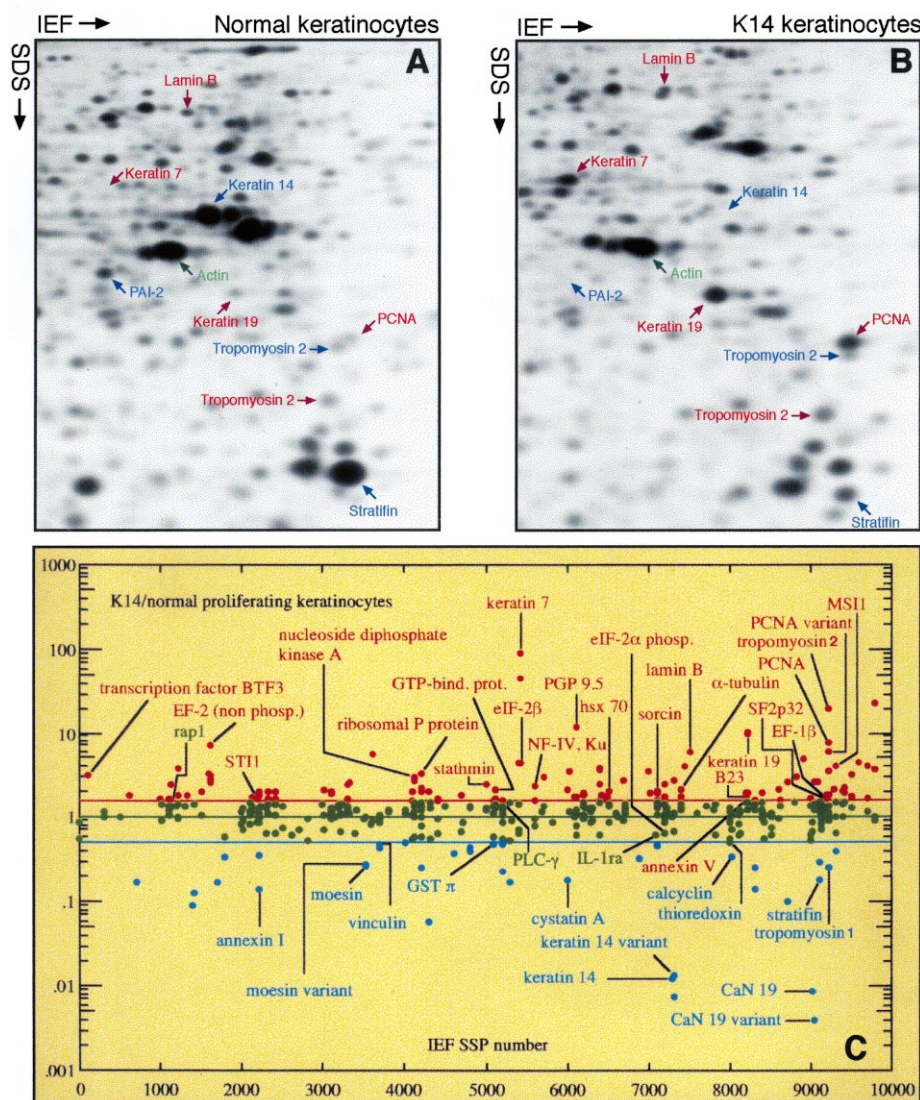
Fig. 2. A and B: IEF 2D gels of non-cultured (A) and SV40-transformed (B) human keratinocytes. Only a fraction of the gels are shown. C: Manual quantitations (scintillation counting) of some of the proteins shown in (A) and (B). Proteins indicated with red are upregulated, those indicated with blue are downregulated, while those indicated with green are unaffected. From Celis and Olsen [68].

weight and provides the highest resolution for protein analysis. Usually, one chooses a condition of interest, for example the addition of serum to non-differentiated human keratinocytes, or compare normal and transformed cells (Fig. 2A,B), and let the cells reveal the global protein response as all detected proteins can be analyzed both qualitatively (post-translational modifications) and quantitatively (relative abundance, co-ordinated expression, Fig. 2C) in relation to each other ([55,68] and references therein; see also http://biobase.dk/cgi-bin/celis).

For many years the 2D PAGE technology relied on the use of carrier ampholytes (amphoteric compounds) to establish the pH gradient, but this technique has proven to be difficult because of the lack of reproducibility created by uncontrollable variations in the batches of ampholytes used to generate the pH gradients. Lately, however, with the introduction of immobilized pH gradients (IPGs) [69,70], which are integral part of the polyacrylamide matrix, it has been possible to obtain focusing patterns that can be easily reproduced by the non-expert. IPGs avoid some of the problems associated

with carrier ampholytes such as cathodic drift and endosmosis, allow a higher loading capacity for micropreparative runs, and provide increased charge resolution when narrow pH gradients (0.03 pH unit/cm) are used ([70,71] and references therein). In our hands, however, carrier ampholytes (3.5–10, Fig. 3A) and broad range IPGs (Fig. 3B) resolve similar number of [$^{35}$S]methionine labelled polypeptides (about 2500) as illustrated with the separation of whole protein extracts from labelled human keratinocytes [72]. It has been proposed that narrow range, overlapping IPG gradients viewed side-by-side may provide a solution to the problem of resolving and depicting the proteome of a given cell type. Recently, however, Corthals and co-workers found this solution unrealistic, as it will require the running of a huge number of gels [71]

Very basic polypeptides have proven difficult to resolve, although both carrier ampholytes (non-equilibrium pH gradient electrophoresis, NEPHGE) [66,73] and IPGs (9–12 and 4–12) ([70] and references therein) have been shown to separate basic proteins.

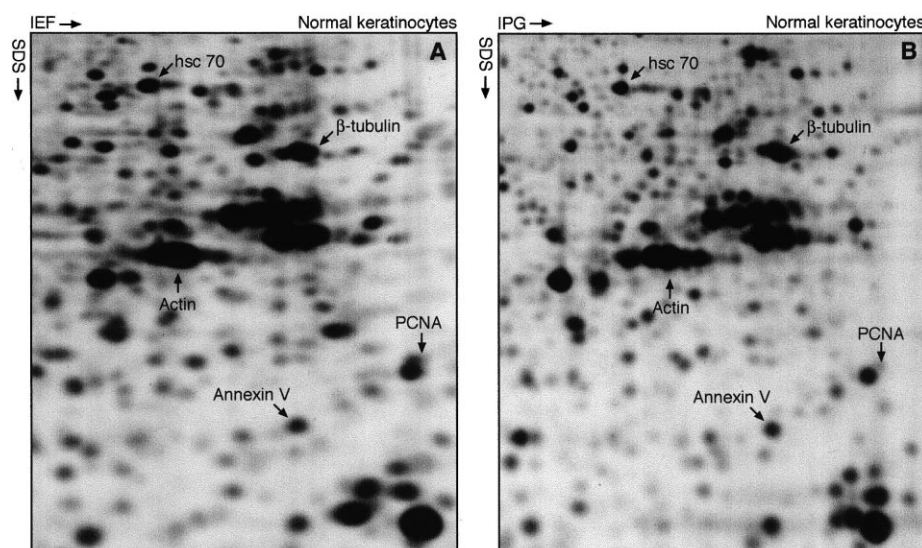One of the most important steps in the 2D PAGE technol-

Fig. 3. Non-cultured normal human keratinocyte proteins separated using (A) carrier ampholytes (3.5–10) and (B) IPGs (3–10). A few proteins are indicated for reference. B is from Bjellqvist et al. [72].

ogy concerns sample preparation as very often some proteins cannot be properly dissolved by the lysis solution originally developed by O'Farrell [65]. Thus, there is pressing need to develop protocols for optimizing sample solubilization. Towards this aim, Rabilloud and co-workers have made use of the high loading capacity of IPGs to resolve membrane proteins for structural analysis, and in doing so have improved their solubility by using a combination of detergents and chaotropes [74,75]. It has been shown that the addition of thiourea, CHAPS and sulfobetain surfactants to the lysis solution containing urea results in a much improved solubilization as well as transfer to the second dimension SDS gel. As far as nuclear proteins are concerned, Görg and colleagues have improved considerably the separation of very basic proteins by first precipitating the samples with acetone prior to solubilization in the lysis solution [70]. The problems associated with the extraction of tissue samples, on the other hand, are much more complex and have not been addressed yet in a systematic fashion.

*3.1.1. Detection.* An important limitation of the 2D PAGE technology is the lack of very sensitive procedures to detect those proteins that are present in very low abundance. In addition, detection procedures are needed that can be applied to a large number of resolved proteins whose abundance may span through seven or eight orders of magnitude. Clearly, the sensitivity of silver nitrate and Coomassie Blue staining is inadequate, and only metabolic labelling with specific isotopes may reveal enough proteins to warrant proteomic projects. Furthermore, the use of phosphor-imaging technology may enhance the sensitivity and linearity of detection. Limitations of the radiolabelling approach include (i) lack of labelling of some proteins due to low turnover, (ii) problems associated with safety regulations and disposal, and (iii) difficulties in obtaining fresh human biopsy material for labelling experiments. Ideally, one would like to have a highly sensitive fluorescence-based protein detection technique able to support all types of studies irrespective of the sample, or the end point of the analysis. Preferably, the dye should not alter the molecular weight and p*I* of the proteins if it is to be added prior to electrophoresis, and should support quantitative studies

involving proteins having extreme differences in their copy numbers. Unfortunately, no such ideal dye is available on the market yet, although Oxford GlycoScience has developed fluorescent IPG–PAGE (http://www.ogs.com/proteome/home.html), a technology not available to the scientific community. Fluorescence compounds such as SYPRO Orange, SYPRO Red and SYPRO Ruby have been used to analyze whole protein lysates from bacterial and mammalian cells, but their sensitivity (1–2 ng) is slightly lower than that of silver nitrate [76,77]. Some advantages over silver staining include short staining time and the fact that the gels do not need to be fixed prior to staining. In addition, little or no destaining is required.

For low abundance proteins of known identity, detection does not pose a problem as Celis and co-workers have shown that 2D PAGE immunoblotting in combination with enhanced chemiluminescence (ECL) can detect as little as 100–500 molecules per cell in unfractionated cellular extracts [60].

*3.1.2. Quantitation.* Even though there are several tools available for the quantitation of protein spots, there is at present no available procedure for quantitating all of the proteins resolved in a complex mixture. Part of the problem lies in the large dynamic range of protein expression, lack of resolution, post-translational modifications, staining behavior of the proteins, as well as in the fact that many abundant proteins streak over less abundant components interfering with the measurements. At present, fluorescent technology seems to be way ahead; as with the fluorescence stain Sypro Ruby there is a linear response with respect to the sample amount over a wide range of abundance ([77] and references therein). Quantitative fluorescence measurements can be performed with CCD-camera based systems as well as with laser scanner systems ([77] and references therein). In some cases, radiolabelling in combination with scintillation counting offers a reasonable alternative for quantitating a small number of proteins [68].

*3.1.3. Identification.* Methods of protein identification have included immunoblotting [78,79], Edman peptide sequencing ([80,81] and references therein), amino acid composition [82,83], and more recently the use of matrix-assisted
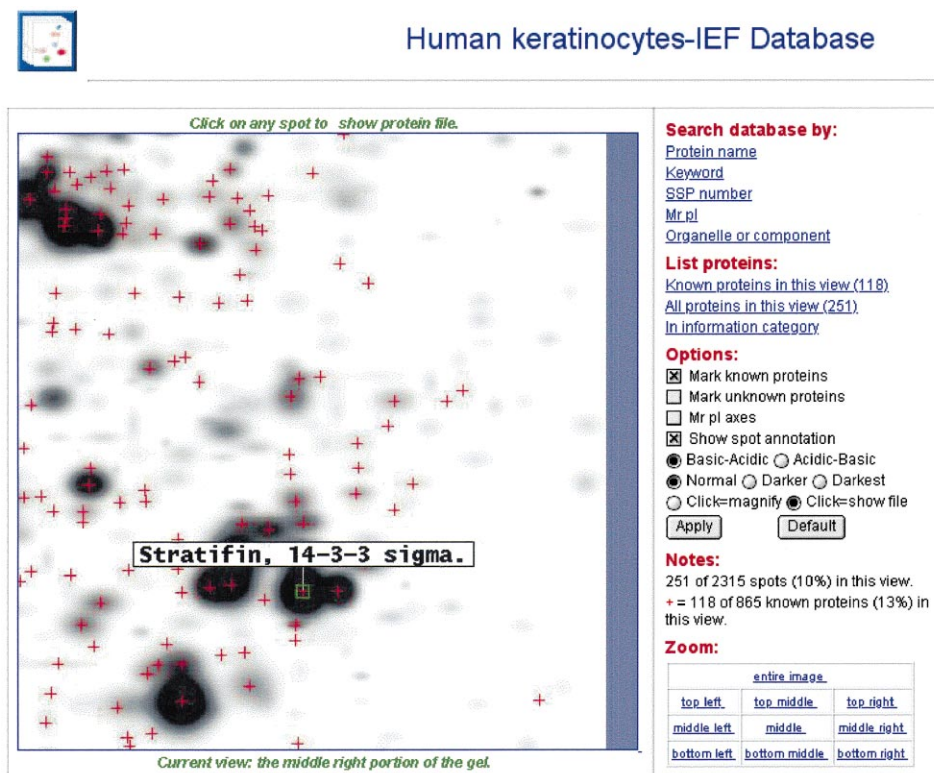
Fig. 4. Fraction of the master synthetic image of human keratinocyte proteins separated by IEF 2D PAGE as depicted in the World Wide Web (http//biobase.dk/cgi-bin/celis). Proteins flagged with a red cross correspond to known proteins. By clicking on any spot it is possible to obtain a file containing information about the protein as well as links to other sites in the World Wide Web.

laser desorption/ionization (MALDI) mass spectrometry [84] and electrospray ionization (ESI) [85]. The latter techniques, which rely on the comparison of peptide mass fingerprints, are fast and require only picomol amounts of proteins. Limited peptide sequencing can be performed using tandem mass spectrometry or post-source decay, but it it is still not possible to make N- and C-terminal sequence identification. For a review on Edman peptide sequencing see reference [81]. Andersen and Mann review protein identification by mass spectrometry elsewhere in this issue.

### 3.2. Making comprehensive 2D PAGE databases

Advances in hardware and software development (Elsie, Gellab, Melanie, Quest, Tycho and Kepler) for scanning and image analysis of 2D gels, as well as the development of sensitive tools for rapid protein identification, have catalyzed the establishment of comprehensive proteomic 2D PAGE databases [64,86]. These databases aim at interfacing protein information with forthcoming DNA mapping and sequence data from genome projects, and offer a global approach to the study of gene expression both in health and disease (http://biobase.dk/cgi-bin/celis; http://expasy.hcuge.ch/sprot/sprot-top.html) ([10,55] and references therein). Besides annotating genomes, these databases are expected to address problems that cannot be approached by DNA analysis, namely, relative abundance of the protein product, post-translational modifications, subcellular localization, turnover, interaction with other proteins as well as functional aspects. The first prototype databases were built by Celis and Bravo [87], who pioneered the use of protein identification techniques to establish comprehensive 2D PAGE databases [88].

Fig. 4 shows a region of the synthetic master 2D PAGE image (isoelectric focusing, IEF) of [$^{35}$S]methionine labelled proteins from fresh non-cultured human keratinocytes as depicted in the World Wide Web (http://biobase.dk/cgi-bin/celis). Proteins flagged with a red cross correspond to known polypeptides that have been identified by one or a combination of techniques that include (i) 2D gel immunoblotting using specific antibodies and the ECL detection procedure, (ii) microsequencing of Coomassie brilliant Blue stained proteins [89], (iii) mass spectrometry of tryptic peptides [90] and (iv) comigration with known human proteins (individual proteins and organelle components) and (v) overlay techniques [59] and (vi) transient expression in mammalian cells [91].

To date, 1237 polypeptides have been identified in this database (IEF and NEPHGE) of the 3159 that have been resolved and catalogued. In addition to [$^{35}$S]methionine labelled proteins, the database contains a few polypeptides that lack methionine, but are revealed by silver staining, Coomassie brilliant Blue or by labelling with a mixture of 16 [$^{14}$C]amino acids. It should be stressed that some proteins migrate both in IEF and NEPHGE gels (for example, α-enolase, triosephosphate isomerase and elongation factor 2) and they may serve as landmarks to align the gels and as references to normalize quantitations in both pH directions.

Information gathered on any given polypeptide, known or unknown, can be retrieved by clicking on the corresponding spot, in this case 14-3-3 sigma, also known as stratifin (Fig. 4) [92]. A file containing all of the information entered for this particular protein, mostly obtained from experiments performed in our laboratory, is shown in Fig. 5 (only a fraction of the file is presented). Files for known proteins contain links

## Spot information: SSP 9109

Stratifin, 14-3-3 sigma.; Mr = 30.0 kDa ; pI = 4.4

For more information, see also MEDLINE or the gopher servers GenBank, Swiss-Prot, PIR, and PDB
Note: Netscape may not return from the Gopher Menu page using 'Go Back'.
Use the Window menu History command to return from the Gopher search.

| Category | Description | Annotation Entry | Link |
|---|---|---|---|
| genecards | Link to GeneCards. | | GeneCards |
| omim | Link to OMIM. | | OMIM |
| prosite | Link to Prosite. | | PROSITE |
| swissprot | Link to Swiss-Prot. | | SWISS-PROT |
| protname | Protein Name. | Stratifin, 14-3-3 sigma. | |
| bfgf | Keratinocyte Proteins Affected by bFGF. | Not affected. J.E.Celis, unpublished observation. | |
| cDNA | | H.Leffers et al., J. Mol. Biol. 231, 982, 1993. | |
| cellulloc | Cellular Localization. | Partially externalized. | |
| levelsk14 | Protein Levels in SV40 Transformed K14 Keratinocytes. | Ratio transformed/normal proliferating = 0.19. J.E.Celis and E.Olsen, Electrophoresis 15, 309, 1994. | |
| levfehutis | Radioactive Levels in Fetal Human Tissues. | Adrenal glands = +; brain = -; cerebellum = -; diaphragm = -; ear = +; eye = -; hypophysis = + (low); kidney = +; large intestine = +; liver = -; lung = + (low); meninges = -; mesonephric tissue = -; pancreas = -; skeletal muscle = -; skin = + (very high); small intestine = +; spleen = -; stomach = +; submandibular gland = + (low); thymus = +; tongue = +;umbilical cord = -; ureter = +. J.E.Celis and A.Celis, unpublished observations. | Normal Histology |
| paraminseq | Patial Amino Acid Sequences. | NLLSVAYK. Residues 43-50; VFYLK. Residues 121-125. H.H.Rasmussen et al., Electrophoresis 13, 960, 1992. | |
| percentotprot | Percentage of Total 14C-Labeled Protein Recovered From IEF and NEPHGE Gels. | 0.799%. J.E.Celis and M.Ostergaard, unpublished observation. | |
| pma | Keratinocyte Proteins Affected by PMA. | Downregulated. J.E.Celis, unpublished observation. | |
| protaffpsor | Proteins Affected in Psoriatic Skin. | Not affected. J.E.Celis et al., unpublished observation. | Psoriasis Foundation |
| retinoicacid | Keratinocyte Proteins Affected by Retinoic Acid. | Not affected. J.E.Celis et al., unpublished observation. | |
| search | Protein Name. | Stratifin, 14-3-3 sigma. | |

Fig. 5. Files for the 14-3-3 sigma protein available in the Internet version of the human keratinocyte database. Only a few entries are shown.

to a subset of Medline (http://www.ncbi.nlm.nih.gov/ PubMed/), Swiss-Prot (http://expasy.hcuge.ch/sprot/sprot-top. html) and PDB (http://www.embl-heidelberg.de/pdb/). Other links include OMIM (http://www.ncbi.nlm.nih.gov/ Omim/), GeneCards (http://bioinformatics.weizmann.ac.il/cards), Uni-Gene (http://www.ncbi.nlm.nih.gov/UniGene/index.html) and other Web sites such as CySPID (cytoskeletal protein database; http://paella.med.yale.edu/∼panzer/cytoskdb/index. html), metabolic pathways (compiled by KEGG; http:// www.genome.ad.jp/kegg/), the cytokine explorer (http:// kbot.mig.missouri.edu:443/cytokines/explorer.html), histology images (http://biosun.biobase. dk/∼pdi/jecelis/micrographs. html), etc. In the future, as new databases and related Web sites become available, it will be possible to navigate throughout various databases containing complementary information (i.e. nucleic acid and protein sequence, genome mapping, diseases, protein structure, post-translational modifications, antibodies, signalling pathways, histology, etc.). Clearly, databases allow easy access to a large body of data: once a protein is identified in given databases, all of the information accumulated can be easily retrieved and made available to the researcher.

Today, 188 information categories are available in the World Wide Web version of the keratinocyte database. These include cellular localization, pathways, proteins affected in psoriatic keratinocytes, proteins expressed in normal urothelium, keratinocytes, fibroblasts, HeLa cells and bladder carcinoma cell line RT4, levels in fetal human tissues, partial amino acid sequences, abundance, cytoskeletal proteins, calcium-binding proteins, annexins, chaperonins, heat shock proteins, etc.

Functions to query the databases include search by name,

protein number or keywords (Fig. 4), molecular weight and pI, as well as by organelle or cellular component. By clicking on any of the organelles, cellular structures or components it is possible to get a protein list as well as their relative positions on the master image. In addition, one can retrieve a list of all known proteins recorded in the database. Moreover, we provide protocols and videos of preparative steps that can be used to reproduce the data displayed as well as a gallery of 2D gels and immunoblots.

Aarhus 2D PAGE databases available at http://biobase.dk/ cgi-bin/celis include transitional cell carcinomas (TCCs), squamous cell carcinomas (SCCs), urine, fibroblasts, and mouse kidney cells. Other databases available in the Internet can be found in http://expasy.hcuge.ch/sprot/sprot-top.html; http:// www.harefield.nthames.nhs.uk/nhli/protein; and http://user-page.chemie.fu-berlin.de/∼pleiss/dhzb.html. Also, several proteomic tools for protein identification and characterization, primary structure analysis, secondary structure prediction, tertiary structure and DNA translation into protein are available at the EXPASY proteomic server (http://www.expasy.ch/www/tools).

### 3.3. Applications

To date, there have been thousands of reports illustrating the usefulness of the 2D PAGE and proteomic technologies in many areas of biology. Because of space limitations, however, only a few of the application areas are highlighted below:

*3.3.1. Cancer.* A great deal of research has been devoted to the elucidation of the pathways that control cell proliferation in normal cells, and hence, the determination of the means by which alterations of these pathways lead to abnormal growth characteristics and/or neoplastic transformation

and cancer. Most studies have focused on oncogenes, tumor suppressors, cell cycle regulated proteins and signal transduction molecules in various cell types of different species [93–96], but only in a few cases there have been systematic attempts to analyze the protein phenotype of pairs of normal and transformed cell types using a proteomic approach [60,68,97–99]. So far, only very few studies have made use of biopsy material due to problems related to the cell heterogeneity.

Already in 1982, Celis and co-workers started a proteomic approach to the study of cell transformation using cloned cell lines [68,97,98]. Their results showed that transformation resulted in the abnormal expression of normal genes, rather than in the expression of new ones [98]. In addition, their studies raised a word of caution concerning the widespread use of protein information derived from studies of different cell types from various species. Today, we are well aware that cultured cells undergo important changes when placed in culture due to different environmental factors and growth conditions [100] and accordingly, current efforts using the proteomic approach are being directed to the study of non-cultured cells and/or tissue biopsies. Among the cancer projects currently underway, those centered on leukemia and hematological malignancies, breast cancer, colorectal cancer and bladder cancer are briefly mentioned below.

*3.3.1.1. Leukemia and hematological malignancies.* Studies by Hanash and colleagues on childhood leukemia and other hematological malignancies have yielded so far several markers that include Op18, also known as stathmin, an oncoprotein that has been implicated in signal transduction [101,102]. In childhood leukemia, phosphorylation of Op18 was shown to correlate with a high content of cells in the S-phase suggesting a role in proliferation. The group also identified nm23-H1 (nucleoside diphosphate kinase A), a 19 kDa protein that is upregulated in normal lymphocytes treated with mitogens as well as in leukemia cells from patients with acute leukemia [103].

*3.3.1.2. Breast cancer.* Systematic studies of clinical breast tumors of different histopathological types by Franzen and co-workers [104–106] have revealed several proteins, including PCNA, hsp60, hsp90 and calreticulin that are highly deregulated in invasive carcinomas and that may serve as prognostic markers. These studies have made use of fresh clinical tumor tissues of different subtypes and have paid special attention to sample preparation.

*3.3.1.3. Colorectal cancer.* Studies of Jungblut and colleagues [107–109] on sets of macroscopically normal colon mucosa and colorectal carcinomas have revealed several proteins that are deregulated in the tumors. Downregulated proteins included the liver fatty acid binding protein, the smooth muscle protein 22-α, and cyclooxygenase 2. Upregulated proteins included the heat shock protein 70 as well as several members of the S-100 family of calcium-binding proteins (S-100A9, S-100 A8, S100A11 and S-100A6). Some of these findings have been confirmed by immunohistochemical studies [108].

*3.3.1.4. Bladder cancer.* Celis and colleagues [110–112] have explored the possibility of using proteome expression profiles of bladder tumors as fingerprints to subclassify histopathological types, and as a starting point for searching for protein markers that may form the basis for diagnosis, prognosis and treatment. To achieve these goals they have analyzed the proteome expression profiles of hundreds of fresh tumors as well as random biopsies and cystectomies [110–112], and have established TCC and SCC proteomic databases that may provide a solid infrastructure to support future studies ([113]; http://biobase.dk/cgi-bin/celis). In the long run, a practical goal of these studies is to identify a complete set of protein biomarkers that may be useful to classify histopathological grades, and that will provide with specific probes for the objective diagnosis, prognosis and treatment of these lesions. So far, these studies have revealed markers for TCC progression [110], a marker in the urine of patients bearing SCCs [114,115], and have led to the development of a novel strategy for identifying premalignant squamous lesions [112]. The approach makes use first of proteomic technologies to reveal and identify proteins that are differentially expressed in pure SCCs and normal urothelium. Thereafter, specific antibodies against the differentially expressed proteins are used to immunostain serial cryostat sections of biopsies (immunowalking) obtained from SCC patients that have undergone removal of the bladder due to invasive disease (cystectomy). Since bladder cancer is a field disease [116] – that is large part of the bladder lining is at risk of developing disease – it is expected that the urothelium of these patients may exhibit a spectrum of abnormalities ranging from metaplasia to invasive disease.

*3.3.2. Heart diseases.* Heart failure is among the leading causes of mortality in the Western Hemisphere and therefore, efforts are being devoted to the elucidation of the molecular events leading to cardiac dysfunction ([117] and references therein). So far, research on dilated cardiomyopathy (DCM) [117–119], has revealed that approximately 100 proteins are deregulated, mostly downregulated, in DCM as compared to their normal counterparts. These include cytoskeletal and myofibrillar proteins, polypeptides associated with mitochondria and involved in energy production, as well as proteins associated with the stress response. These studies have been expedited by the establishment of proteomic 2D PAGE databases of the human heart (ventricle and atrium) [120]

In addition to the global analysis of protein expression patterns in human heart diseases, cardiac antigen expression following cardiac transplantation has also been studied using techniques from proteomics. For example, using 2D PAGE (cardiac proteins) in combination with Western immunoblotting (patient sera), it has been possible to identify antigens that react with autoantibodies present both in DCM [121,122] and myocarditis [123]. In this way, antigens associated with the antibody response that may be involved in acute or chronic organ rejection have been characterized.

Proteomic studies using animal models of heart disease have also been carried out in an effort to unravel the molecular events leading to cardiac disease. Recently, two different studies involving large animals, pace-induced heart failure in the dog [124] and bovine DCM [125], yielded similar results as those observed in human DCM. Interestingly, the most striking result in the bovine DCM study was the finding of a seven-fold decrease in the expression level of ubiquitin C-terminal hydrolase, as inappropriate ubiquination of proteins has been suggested as an etiologic factor in heart failure [126].

*3.3.3. Toxicology.* Changes in the environment as well as the growing interest of the pharmaceutical industry have stimulated the development of novel testing approaches based on the recent technical advances both in genomics and proteomics. Pioneering studies by Anderson and Anderson [127]
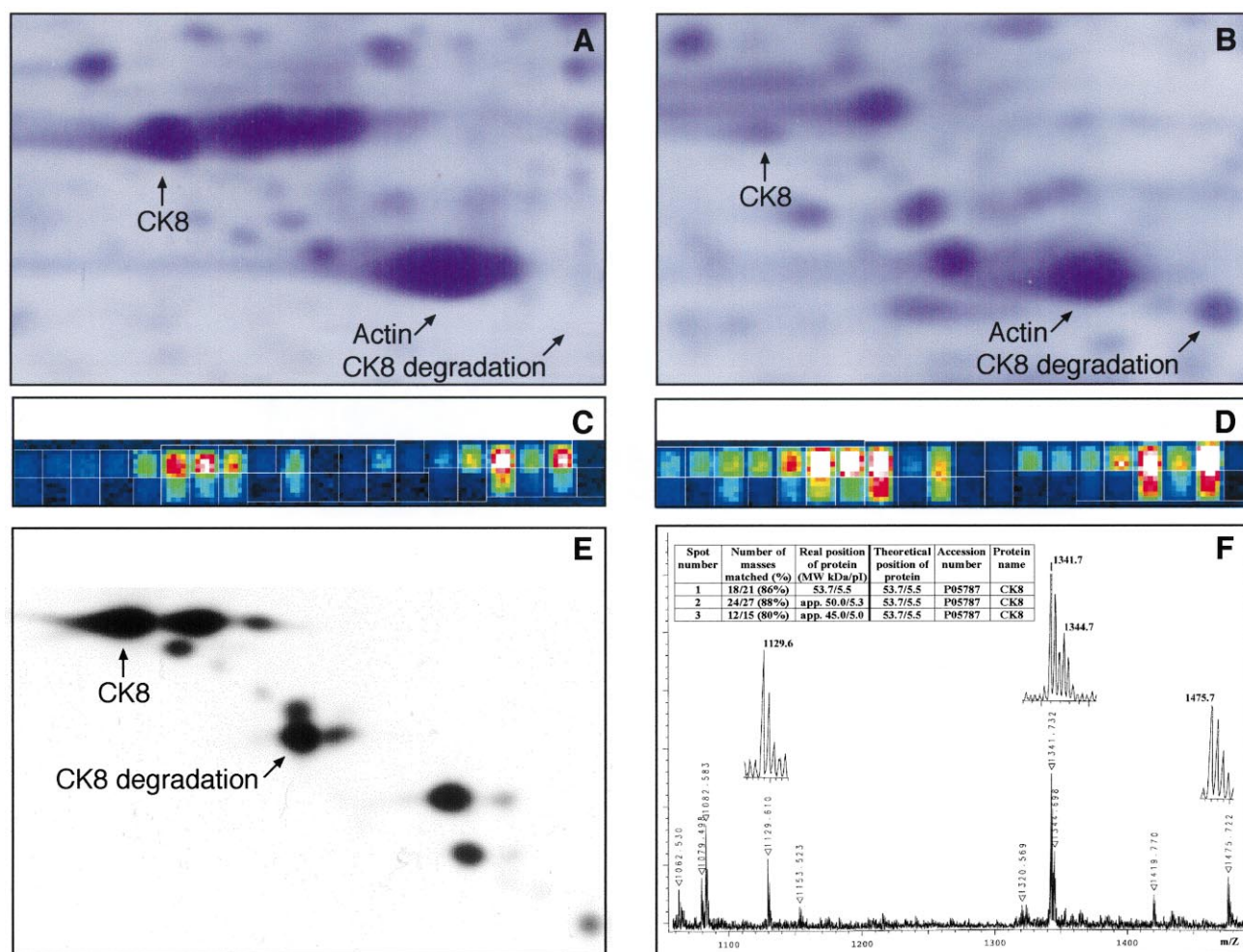
| Spot number | Number of masses matched (%) | Real position of protein (MW kDa/pI) | Theoretical position of protein | Accession number | Protein name |
|---|---|---|---|---|---|
| 1 | 18/21 (86%) | 53.7/5.5 | 53.7/5.5 | P05787 | CK8 |
| 2 | 24/27 (88%) | app. 50.0/5.3 | 53.7/5.5 | P05787 | CK8 |
| 3 | 12/15 (80%) | app. 45.0/5.0 | 53.7/5.5 | P05787 | CK8 |

Fig. 6. 2D PAGE (A and B) and chip data (C and D) on the expression of keratin 8 in a non-invasive (grade II, Ta) and an invasive TCC (grade III, T1). C and D: The top row shows reaction with perfect match probes, the lower row shows reaction with mismatch control probes. E: Immunoblot of the proteins resolved in A reacted with antibodies against keratin 8. The position of keratin 8 and of one of its degradation products is indicated for reference. F: MALDI-TOF peptide fingerprint of the keratin 8 degradation product indicated in A and E.

have highlighted the usefulness and potential of the proteomic approach to identify quantitative changes in rat liver expression profiles associated with toxicity of drugs and other xenobiotics. The data, which are being systematically stored in the rodent molecular effects database, are expected to yield important information as to the molecular mechanisms underlying toxic responses. Likewise, the potential of proteomics have been recently exemplified in studies of glomerular nephrotoxicity in rats [128], and of stimulated occupational jet fuel exposure in mice lung [129]. In particular, the studies of Steiner's group have shown a remarkable correlation between decreased levels of calbindin D-28, urinary calcium wasting in the urine, and intratubular corticomedullary calcifications in the kidney of rats and human treated with cyclosporin A [130].

In the future, proteomics in a high-throughput mode is expected to have a major impact in the pre-clinical safety testing of drugs. These studies will be facilitated by the establishment of 2D PAGE databases of frequent target tissues (kidney, liver) as well as of cell lines and fluids.

*3.3.4. Neurological disorders.* The Creutzfeld–Jacob disease (CJD) has been the subject of intensive analysis using proteomics. These studies have led to the identification of two members of the 14-3-3 family of proteins in the cerebral spinal fluid (CSF) of CJD patients [131]. The presence of these proteins in the CSF has been used to differentiate CJD from other dementia both with high sensitivity and specificity [132,133]. These proteins, however, are present in the CSF of patients suffering from other neurological disorders not involving dementia, limiting its clinical value [134,135].

*3.4. Conclusions*

Today, there is no technology in sight that matches the resolving power of 2D PAGE, a technique that will continue to enjoy a central position in proteomic projects for some time in the near future. There is considerable room for improvement, however, in particular as far as sample preparation and solubility, choice of pH gradient and detection methods are concerned. Also, we need to improve the separation of very basic as well as very low and high molecular weight polypeptides. In general, one expects researchers to first use wide IPG gradients to obtain an overview of the proteome profiles, and then proceed with a more detailed analysis using narrow pH gradients, which provide higher resolution and sample loading, thus increasing the possibility of visualizing the lesser abundant proteins. The latter can be facilitated by the avail-

ability of specific antibodies, as well as by the use of extraction procedures and subcellular fractionation methods currently at hand [136–138].

There are still many additional challenges, however, that must be addressed before a complete Human Proteome Project can be implemented [139]. These include: automation to allow high-throughput sample analysis [140], improved quantitation capabilities, better instrumentation and software for peptide sequencing using mass spectrometry, more sophisticated image analysis systems to support gel comparisons and databasing as well as improved bioinformatic capabilities overall [12]. In addition, we need to deal with the problem of tissue cell heterogeneity as more and more proteomic projects will make use of biopsy material in the future.

## 4. Transcript and protein levels: DNA microarrays and proteomics applied to the same samples

As mentioned in Section 1, both DNA microarrays and proteomics are complementary technologies. To date, however, there have been only a limited number of studies in which both technologies have been compared by applying them to the same sample [141,142]. Notably, the pioneer studies of Anderson and Seilhamer [52] showed that there is not a good correlation between mRNA and protein levels in human liver, implying that gene-based expression data may be of limited value in the process of drug discovery. The study, which compared the levels of 19 gene products, yielded a correlation coefficient of 0.48 between mRNA and protein abundance, a value that is half way between perfect and no correlation.

Recently, Ørntoft et al. (manuscript in preparation) carried out a microarray and proteomic study of bladder cancer in which they compared the transcript and protein expression levels of pairs of non-invasive and invasive low grade fresh TCCs. Even though they could only compare the levels of about 40 well-resolved and focused abundant proteins, it was clear that in most cases there was a good correlation between transcript and protein levels. Only in a few cases they found discrepancies, and in some of those instances they could not eliminate the possibility that this was due to messenger stability, post-transcriptional splicing, post-translational modifications, protein focusing problems, degradation, as well as the choice of methods used to assess protein expression levels (staining versus radiolabelling). For example, in one tumor pair they found that the levels of keratin 8 transcripts were much higher in the invasive tumor (compare Fig. 6C and D), while the protein levels were much lower (compare Fig. 6A and B). Immunoblotting analysis using keratin 8 specific antibodies revealed that the discrepancy was due to degradation, as several related products of lower apparent molecular weights and more acidic p$I$s, could be visualized (Fig. 6E). The identity of one of the crossreacting peptides (indicated with arrows in Fig. 6E) to keratin 8 was further confirmed by MALDI-time-of-flight (TOF) (Fig. 6F). From these studies it was clear that when comparing mRNA and protein levels there are other factors that need to be taken into consideration when interpreting the data.

When comparing transcripts and protein expression profiles of matched sample pairs one often gets the impression that there are more changes in the abundance of the mRNA tran-

scripts as compared to the proteins. Considering that the current 2D PAGE technology depicts mainly the more abundant proteins, it would seem possible that most of the changes affecting protein levels may involve low abundance polypeptides.

## 5. Gene expression profiling techniques: perspectives

Novel and powerful techniques are now available to analyze the global gene expression patterns of cultured cells and tissues obtained from normal and diseased subjects. Each of these technologies has its own advantages and limitations, but in combination they should provide us with a detailed gene expression phenotype at both the transcription and translation level. A major challenge in the near future will be to define a base line for the normal gene expression phenotype of a given cell type, tissue or body fluid. This is not a trivial task, however, as it will require the analysis of hundreds or even thousands of samples.

Besides improvements on the individual techniques themselves (see above), there are still major limitations that must be addressed before these technologies can provide the expected outcome in molecular medicine. These include: (i) technical problems associated with the analysis of expression profiles derived from tissues that are composed of different cell types, (ii) lack of procedures for identifying targets that lie in the pathway of disease, and (iii) need for bioinformatics tools for rapidly assessing the function of the putative targets. The latter is of paramount importance to the pharmaceutical industry as the identification of disease deregulated targets alone is not sufficient to start a costly drug screening process.

As far as tissue heterogeneity is concerned, the recent advent of LCM holds great promise as with this technique it is possible to isolate specific populations of cells from a tissue section using direct microscopic observation [29]. However, even though the technique has been used for RNA analysis it is still not ready for most proteomic projects as the number of cells that can be obtained is too small to generate reasonable protein profiles in terms of the number of proteins that can be visualized [143].

As DNA microarrays and proteomics generate more data in the future it will become a matter of priority to develop simple and rapid strategies to validate the vast amount of information that will be generated, particularly in tissue biopsies. This we believe can be accomplished in part by making use of specific antibodies in combination with immunohistochemistry [112]. At present, there is no technology at hand that may allow us to prepare antibodies at will, although phage antibody libraries [14] show much promise.

Finally, we would like to emphasize that biology in this Millennium will be characterized by the study of complex biological phenomena. DNA microarrays and proteomics are just some of the technologies of functional genomics, and only their integration may allow us to tackle the great complexity underlying biological processes.

# References

[1] Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matize, T.C., McKusick, K.B., Beckmann, J.S., Bentolila, S., Bihoreau, M., Birren, B.B., Browne, J., Butler, A., Castle, A.B., Chiannilkulchai, N., Clee, C., Day, P.J., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C. and Bentley, D.R. et al. (1998) Science 282, 744–746.
[2] Collins, A. (2000) Hum. Hered. 50, 76–84.
[3] Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Science 20, 467–470.
[4] Gerhold, D., Rushmore, T. and Caskey, C.T. (1999) Trends Biochem. Sci. 24, 168–173.
[5] Lennon, G.G. (2000) Drug Discov. Today 5, 59–66.
[6] Liang, P. and Pardee, A.B. (1992) Science 14, 967–971.
[7] Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Science 270, 484–487.
[8] Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. (1997) Science 23, 1268–1272.
[9] Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W. and Vogelstein, B. (1997) Nature 18, 300–305.
[10] Celis, J.E., Østergaard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H. and Gromov, P. (1998) FEBS Lett. 430, 64–72.
[11] Appel, R.D. and Dunn, M.J. (Eds.), (1999) Electrophoresis 20, 3481–3686.
[12] Dunn, M.J. (Ed.), (2000) Electrophoresis 21, 1035–1236.
[13] Shashikant, C.S., Carr, J.L., Bhargava, J., Bentley, K.L. and Ruddle, F.H. (1998) Gene 223, 9–20.
[14] McCafferty, J. Hoogenboom, H.R. and Chiswell, D. J. (Eds.) (1996) Antibody Engineering, Oxford University Press, Oxford.
[15] Nelson, R.W. and Nedelkov, D. (2000) Electrophoresis 21, 1155–1163.
[16] Wallach, D., Boldin, M.P., Kovalenko, A.V., Malinin, N.L., Mett, I.L. and Camonis, J.H. (1998) Curr. Opin. Immunol. 10, 131–136.
[17] Skolknick, J. and Fetrow, J.S. (2000) Trends Biotechnol. 18, 34–39.
[18] Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. (1996) Science 274, 610–614.
[19] Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.. and Lander, E.S. (1998) Science 280, 1077–1082.
[20] Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P. (1994) Proc. Natl. Acad. Sci. USA 91, 5022–5026.
[21] Lipshutz, R.J., Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R. and Fodor, S.P. (1995) Biotechniques 19, 442–447.
[22] Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996) Nat. Biotechnol. 14, 1675–1680.
[23] Takenaka, M., Imai, E., Nagasawa, Y., Matsuoka, Y., Moriyama, T., Kaneko, T., Hori, M., Kawamoto, S. and Okubo, K. (2000) Kidney Int. 57, 19–24.
[24] Beier, M. and Hoheizel, J.D. (1999) Nucleic Acids Res. 27, 1970–1977.
[25] Mahadevappa, M. and Warrington, J.A. (1999) Nat. Biotechnol. 17, 1134–1136.
[26] Perou, C.M., Jeffrey, S.S., Van De, R.M., Rees, C.A., Eizen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C., Lashkari, D., Shalon, D., Brown, P.O. and Botstein, D. (1999) Proc. Natl. Acad. Sci. USA 96, 9212–9217.
[27] Sgroi, D.C., Teng, S., Robinson, G., LeVangie, R., Hudson Jr., J.R. and Elkahloun, A.G. (1999) Cancer Res. 59, 5656–5661.
[28] Luo, L., Salunga, R.C., Guo, H., Bittner, A., Joy, K.C., Galindo, J.E., Xiao, H., Rogers, K.E., Wan, J.S., Jackson, M.R. and Erlander, M.G. (1999) Nat. Med. 5, 117–122.
[29] Bonner, R.F., Emmert-Buck, M., Cole, K., Pohida, T., Chuaqui, R., Goldstein, S. and Liotta, L.A. (1997) Science 278, 1481–1483.
[30] Everitt, B.S. (1998) Cluster Analysis, Arnold, London.
[31] Fowler, J., Cohen, L. and Jarvis, P. (1998) pp. 210–234, John Wiley and Sons, Chichester.
[32] Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999) Proc. Natl. Acad. Sci. USA 96, 2907–2912.
[33] Törönen, P., Kolehmainen, M., Wong, G. and Castrén, E. (1999) FEBS Lett. 451, 142–146.
[34] Hartigan, J. (1975) Clustering Algoritms, John Willey and Sons, New York.
[35] Eizen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Proc. Natl. Acad. Sci. USA 95, 14863–14868.
[36] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. and Lander, E.S. (1999) Science 286, 531–537.
[37] Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M.J. and Haussler, D. (2000) Proc. Natl. Acad. Sci. USA 97, 262–267.
[38] Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. (1999) Proc. Natl. Acad. Sci. USA 96, 6745–6750.
[39] Hiltunen, M.O., Niemi, M. and Yla-Herttuala, S. (1999) Curr. Opin. Lipidol. 10, 515–519.
[40] McCaffrey, T.A., Fu, C., Du, B., Eksinar, S., Kent, K.C., Bush, H.Jr., Kreiger, K., Rosengart, T., Cybulsky, M.I., Silverman, E.S. and Collins, T. (2000) J. Clin. Invest 105, 653–662.
[41] Ross, D.T., Scherf, U., Eizen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van De, R.M., Waltham, M., Pergamenschikov, A., Lee, J.C., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D. and Brown, P.O. (2000) Nat. Genet. 24, 227–235.
[42] Gonzalez, P., Epstein, D.L. and Borras, T. (2000) Invest. Ophthalmol. Vis. Sci. 41, 352–361.
[43] Scarlato, M., Beesley, J. and Pleasure, D. (2000) J. Neurosci. Res. 59, 430–435.
[44] Der, S.D., Zhou, A., Williams, B.R. and Silverman, R.H. (1998) Proc. Natl. Acad. Sci. USA 95, 15623–15628.
[45] Zhu, H., Cong, J.P., Mamtora, G., Gingeras, T. and Shenk, T. (1998) Proc. Natl. Acad. Sci. USA 95, 14470–14475.
[46] Khan, J., Bittner, M.L., Saal, L.H., Teichmann, U., Azorsa, D.O., Gooden, G.C., Pavan, W.J., Trent, J.M. and Meltzer, P.S. (1999) Proc. Natl. Acad. Sci. USA 96, 13264–13269.
[47] Kaminski, N., Allard, J.D., Pittet, J.F., Zuo, F., Griffiths, M.J., Morris, D., Huang, X., Sheppard, D. and Heller, R.A. (2000) Proc. Natl. Acad. Sci. USA 97, 1778–1783.
[48] Rockett, J.C. and Dix, D.J. (2000) Xenobiotica 30, 155–177.
[49] Bartosiewicz, M., Trounstine, M., Barker, D., Johnston, R. and Buckpitt, A. (2000) Arch. Biochem. Biophys. 376, 66–73.
[50] Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T., Scudiero, D.A., Eizen, M.B., Sausville, E.A., Pommier, Y., Botstein, D., Brown, P.O. and Weinstein, J.N. (2000) Nat. Genet. 24, 236–244.
[51] Alizadeh, A.A., Eizen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J.Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weizenburger, D.D., Armitage, J.O., Warnke, R. and Staudt, L.M. (2000) Nature 403, 503–511.
[52] Anderson, L. and Seilhamer, J. (1997) Electrophoresis 18, 533–537.
[53] Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F. and Williams, K.L. (1996) Biotechnol. Genet. Eng. Rev. 13, 19–50.
[54] Blackstock, W.P. (1999) Trends. Biotechnol. 17, 121–127.
[55] Celis, J.E., Rasmussen, H.H., Leffers, H., Madsen, P., Honore, B., Gesser, B., Dejgaard, K. and Vandekerckhove, J. (1991) FASEB J. 5, 2200–2208.
[56] Celis, J.E. (Ed.), (1996) Electrophoresis 17, 1653–1798.
[57] Humphery-Smith, I. (Ed.), (1997) Electrophoresis 18, 1205–1498.
[58] Huber, L.A. (Ed.), (1997) Electrophoresis 18, 2505–2700.
[59] Celis, J.E. (1999) Curr. Opin. Biotechnol. 10, 16–21.
[60] Celis, J.E., Rasmussen, H.H., Gromov, P., Olsen, E., Madsen, P., Honoré, B., Dejgaard, K., Vorum, H., Kristensen, D.B., Øster-

gaard, M., Jensen, N.A., Celis, A., Basse, B., Lauridsen, J., Ratx, G.P., Walbum, E., Andersen, I., Puype, M., Van Damme, J. and Vandekerckhove, J. (1995) Electrophoresis 16, 2177–2240.

[61] Gooley, A.A and Packer, N.H.. (1997). in: Proteome research: New Frontiers in Functional Genomics (Principles and Practice) (Wilkins, M.R., Williams, K.L, Appel, R.D and Hochstrasser, D.F., Eds.), Springer Verlag, Berlin.

[62] (1982) Clin. Chem. 28, 737–1092.

[63] Celis, J.E and Bravo, R. (Eds.) (1984) Two-dimensional Gel Electrophoresis of Proteins: Methods and Applications, Academic Press, New York.

[64] Dunn, M.J. (1999) Electrophoresis 20, 639–1122.

[65] O'Farrell, P. (1975) J. Biol. Chem. 250, 4007–4021.

[66] O'Farrell, P.Z., Goodman, H.M. and O'Farrell, P. (1977) Cell 12, 1133–1141.

[67] Klose, J. (1975) Humangenetik 26, 231–243.

[68] Celis, J.E. and Olsen, E. (1994) Electrophoresis 15, 309–344.

[69] Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Görg, A., Westermeier, R. and Postel, W. (1982) J. Biochem. Biophys. Methods 6, 317–339.

[70] Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and Weiss, W. (2000) Electrophoresis 21, 1037–1053.

[71] Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F. and Sanchez, J.C. (2000) Electrophoresis 21, 1104–1115.

[72] Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E. (1994) Electrophoresis 15, 529–539.

[73] Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., Celis, A., Jensen, N.A. and Gromov, P. (1997). in: Cell Biology: A Laboratory Handbook (Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K. and Small, J.V., Eds.), Vol. IV, pp. 375–385, Academic Press, San Diego, CA.

[74] Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J. (1997) Electrophoresis 18, 307–316.

[75] Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J. and Rabilloud, T. (1998) Electrophoresis 19, 1901–1909.

[76] Steinberg, T.H., Haugland, R.P. and Singer, V.L. (1996) Anal. Biochem. 239, 238–245.

[77] Patton, W.F. (2000) Electrophoresis 21, 1123–1144.

[78] Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350–4354.

[79] Celis, J.E., Lauridsen, J.B. and Basse, B. (1997). in: Cell Biology: A Laboratory Handbook (Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K. and Small, J.V., Eds.), Vol. I, pp. 429–437, Academic Press, San Diego, CA.

[80] Rasmussen, H.H., Van Damme, J., Bauw, G., Puype, M., Celis, J.E. and Vandekerckhove, J. (1991) in: Methods in Protein Sequence Analysis, (Jörnvall, H., Höög, J.-O. and Gustavsson, A.-M. Eds.), pp. 103–104, Birkhäuser, Basel.

[81] Gevaert, K. and Vandekerckhove, J. (2000) Electrophoresis 21, 1145–1154.

[82] Shaw, G. (1993) Proc. Natl. Acad. Sci. USA 90, 5138–5142.

[83] Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L. and Hochstrasser, D.F. (1996) Biotechnology (NY) 14, 61–65.

[84] Karas, M. and Hillenkamp, F. (1988) Anal. Chem. 60, 2299–2301.

[85] Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) Science 246, 64–71.

[86] Celis, J.E. (Ed.), (1992) Electrophoresis 13, 861–1062.

[87] Bravo, R. and Celis, J.E. (1982) Clin. Chem. 28, 766–781.

[88] Celis, J.E., Rasmussen, H.H., Madsen, P., Leffers, H., Honore, B., Dejgaard, K., Gesser, B., Olsen, E., Gromov, P., Hoffmann, H.J., Nielsen, M., Celis, A., Basse, B., Lauridsen, J.B., Ratz, G., Nielsen, H., Andersen, A.H., Walbum, E., Kjærgaard, I., Puype, M., Van Damme, J. and Vandekerckhove, J. (1992) Electrophoresis 13, 893–959.

[89] Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Ratz, G.P., Lauridsen, J.B. and Celis, J.E. (1989) Proc. Natl. Acad. Sci. USA 86, 7701–7705.

[90] Rasmussen, H.H., Mörtz, E., Mann, M., Roepstorff, P. and Celis, J.E. (1994) Electrophoresis 15, 406–416.

[91] Madsen, P., Gromov, P. and Celis, J.E. (1995) Electrophoresis 16, 2258–2261.

[92] Leffers, H., Madsen, P., Rasmussen, H.H., Honore, B., Andersen, A.H., Walbum, E., Vandekerckhove, J. and Celis, J.E. (1993) J. Mol. Biol. 20, 982–998.

[93] Kinzler, K.W. and Vogelstein, B. (1996) Cell 87, 159–170.

[94] Hunter, T. (1997) Cell 88, 333–346.

[95] Levine, A.J. (1997) Cell 88, 323–331.

[96] Hanahan, D. and Weinberg, R.A. (2000) Cell 100, 57–70.

[97] Bravo, R. and Celis, J.E. (1980) Exp. Cell. Res. 127, 249–260.

[98] Bravo, R. and Celis, J.E. (1982) Clin. Chem. 28, 949–954.

[99] Wirth, P.J., Luo, L.D., Fujimoto, Y. and Bisgaard, H.C. (1992) Electrophoresis 13, 305–320.

[100] Celis, A., Rasmussen, H.H., Celis, P., Basse, B., Lauridsen, J.B., Ratz, G., Hein, B., Østergaard, M., Wolf, H., Ørntoft, T. and Celis, J.E. (1999) Electrophoresis 20, 355–361.

[101] Strahler, J.R., Lamb, B.J., Ungar, D.R., Fox, D.A. and Hanash, S.M. (1992) Biochem. Biophys. Res. Commun. 185, 197–203.

[102] Melhem, R., Hailat, N., Kuick, R. and Hanash, S.M. (1997) Leukemia 11, 1690–1695.

[103] Keim, D., Hailat, N., Melhem, R., Zhu, X.X., Lascu, I., Veron, M., Strahler, J. and Hanash, S.M. (1992) J. Clin. Invest. 89, 919–924.

[104] Franzen, B., Auer, G., Alaiya, A.A., Eriksson, E., Uryu, K., Hirano, T., Okuzawa, K., Kato, H. and Linder, S. (1996) Int. J. Cancer 69, 408–414.

[105] Franzen, B., Linder, S., Alaiya, A.A., Eriksson, E., Uryu, K., Hirano, T., Okuzawa, K. and Auer, G. (1996) Br. J. Cancer 74, 1632–1638.

[106] Franzen, B., Linder, S., Uryu, K., Alaiya, A.A., Hirano, T., Kato, H. and Auer, G. (1996) Br. J. Cancer 73, 909–913.

[107] Stulik, J., Kovarova, H., Macela, A., Bures, J., Jandik, P., Langr, F., Otto, A. and Thiede, B. (1997) Clin. Chim. Acta 265, 41–55.

[108] Stulik, J., Osterreicher, J., Koupilova, K., Knizek, J., Macela, A., Bures, J., Jandik, P., Langr, F., Dedic, K. and Jungblut, P.R. (1999) Electrophoresis 20, 1047–1054.

[109] Stulik, J., Koupilova, K., Osterreicher, J., Knizek, J., Macela, A., Bures, J., Jandik, P., Langr, F., Dedic, K. and Jungblut, P.R. (1999) Electrophoresis 20, 3638–3646.

[110] Celis, J.E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J.B., Ratz, G.P., Andersen, I., Hein, B., Wolf, H., Ørntoft, T.F. and Rasmussen, H.H. (1996) Cancer Res. 56, 4782–4790.

[111] Østergaard, M., Rasmussen, H.H., Nielsen, H.V., Vorum, H., Ørntoft, T.F., Wolf, H. and Celis, J.E. (1997) Cancer Res. 57, 4111–4117.

[112] Celis, J.E., Celis, P., Østergaard, M., Basse, B., Lauridsen, J.B., Ratz, G., Rasmussen, H.H., Ørntoft, T.F., Hein, B., Wolf, H. and Celis, A. (1999) Cancer Res. 59, 3003–3009.

[113] Celis, J.E., Østergaard, M., Rasmussen, H.H., Gromov, P., Gromova, I., Varmark, H., Palsdottir, H., Magnusson, N., Andersen, I., Basse, B., Lauridsen, J.B., Ratz, G., Wolf, H., Ørntoft, T.F., Celis, P. and Celis, A. (1999) Electrophoresis 20, 300–309.

[114] Celis, J.E., Rasmussen, H.H., Vorum, H., Madsen, P., Honoré, B., Wolf, H. and Ørntoft, T.F. (1996) J. Urol. 155, 2105–2112.

[115] Østergaard, M., Wolf, H., Ørntoft, T.F. and Celis, J.E. (1999) Electrophoresis 20, 349–354.

[116] Slaughter, D.P., Southwick, H.W. and Smejkal, W. (1953) Cancer 6, 963–968.

[117] Dunn, M.J. (2000) Drug Discov. Today 5, 76–84.

[118] Corbett, J.M., Why, H.J., Wheele, rC.H., Richardson, P.J., Archard, L.C., Yacoub, M.H. and Dunn, M.J. (1998) Electrophoresis 19, 2031–2042.

[119] Li, X.P., Pleissner, K.P., Scheler, C., Regitz-Zagrosek, V., Salnikow, J. and Jungblut, P.R. (1999) Electrophoresis 20, 891–897.

[120] Evans, G., Wheeler, C.H., Corbett, J.M. and Dunn, M.J. (1997) 18, 471–479.

[121] Latif, N., Baker, C.S., Dunn, M.J., Rose, M.L., Brady, P. and Yacoub, M.H. (1993) J. Am. Coll. Cardiol. 22, 1378–1384.

[122] Pohlner, K., Portig, I., Pankuweit, S., Lottspeich, F. and Maisch, B. (1997) Am. J. Cardiol. 15, 1040–1045.

[123] Pankuweit, S., Portig, I., Lottspeich, F. and Maisch, B.J. (1997) J. Mol. Cell. Cardiol. 29, 77–84.

[124] Heinke, M.Y., Wheeler, C.H., Yan, J.X., Amin, V., Chang, D., Einstein, R., Dunn, M.J. and dos Remedios, C.G. (1999) Electrophoresis 20, 2086–2093.

[125] Weekes, J., Wheeler, C.H., Yan, J.X., Weil, J., Eschenhagen, T., Scholtysik, G. and Dunn, M.J. (1999) Electrophoresis 20, 898–906.

[126] Field, M.L. and Clark, J.F. (1997) Cardiovasc. Res. 33, 8–12.

[127] Anderson, N.G. and Anderson, N.L. (1996) Electrophoresis 17, 443–453.

[128] Cutler, P., Bell, D.J., Birrell, H.C., Connelly, J.C., Connor, S.C., Holmes, E., Mitchell, B.C., Monte, S.Y., Neville, B.A., Pickford, R., Polley, S., Schneider, K. and Skehel, J.M. (1999) Electrophoresis 20, 3647–3658.

[129] Witzmann, F.A., Bauer, M.D., Fieno, A.M., Grant, R.A., Keough, T.W., Kornguth, S.E., Lacey, M.P., Siegel, F.L., Sun, Y., Wright, L.S., Young, R.S. and Witten, M.L. (1999) Electrophoresis 20, 3659–3669.

[130] Aaicher, L., Wahl, D., Arce, A., Grenet, O. and Steiner, S. (1998) Electrophoresis 19, 1998–2003.

[131] Harrington, M.G., Merril, C.R., Asher, D.M. and Gajdusek, D.C. (1986) N. Engl. J. Med. 315, 279–283.

[132] Zerr, I., Bodemer, M., Gefeller, O., Otto, M., Poser, S., Wiltfang, J., Windl, O., Kretzschmar, H.A. and Weber, T. (1998) Ann. Neurol. 43, 32–40.

[133] Beaudry, P., Cohen, P., Brandel, J.P., Delasnerie-Laupretre, N., Richard, S., Launay, J.M. and Laplanche, J.L. (1999) Dementia Geriatr. Cogn. Disord. 10, 40–46.

[134] Zeidler, M., Stewart, G.E., Barraclough, C.R., Bateman, D.E., Bates, D., Burn, D.J., Colchester, A.C., Durward, W., Fletcher, N.A., Hawkins, S.A., Mackenzie, J.M. and Will, R.G. (1997) Lancet 350, 903–907.

[135] Hill, A.F., Butterworth, R.J., Joiner, S., Jackson, G., Rossor, M.N., Thomas, D.J., Frosh, A., Tolley, N., Bell, J.E., Spencer, M., King, A., Al-Sarraj, S., Ironside, J.W., Lantos, P.L. and Collinge, J. (1999) Lancet 353, 183–189.

[136] Bravo, R., Small, J.V., Fey, S.J., Larsen, P.M. and Celis, J.E. (1982) J. Mol. Biol. 154, 121–143.

[137] Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K. and Small, J.V. (Eds.) (1997), vol. 2, Academic Press, San Diego, CA.

[138] Cordwell, S.J., Nouwens, A.S., Verrills, N.M., Basseal, D.J. and Walsh, B. (2000) Electrophoresis 21, 1094–1103.

[139] Butler, D. (2000) Nature 403, 815–816.

[140] Lopez, M.F. (2000) Electrophoresis 21, 1082–1093.

[141] Anderson, N.L. and Anderson, N. (1998) Electrophoresis 19, 1853–1861.

[142] Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999) Mol. Cell. Biol. 19, 1720–1730.

[143] Banks, R.E., Dunn, M.J., Forbes, M.A., Stanley, A., Pappin, D., Naven, T., Gough, M., Harnden, P. and Selby, P.J. (1999) Electrophoresis 20, 689–700.