Chimia 53 (1999) 57–61
© Neue Schweizerische Chemische Gesellschaft
ISSN 0009–4293

Biosensors in Biomedical Research: Development and Applications of Gene Chips

Ulrich Certa[a]*, Remo Hochstrasser[a], Hanno Langen[a], Martin Buess[b], and Christoph Moroni[b]

Abstract. Nucleic-acid hybridisation techniques are a central tool for the genetic analysis of biological systems. Gene chips are complex arrays of recombinant plasmids or oligonucleotides immobilised on a glass chip of only 1 cm^2. This technology allows, for the first time, the multiparallel expression-analysis of thousands of genes. Gene chips will be indispensable tools for the upcoming analysis of the human genome, once the entire sequence is known.

1. Introduction

Molecular biology is a relatively young branch of the life sciences and it deals mainly with the structure and function of genomes. The obvious and challenging goal is to understand how a multicellular organism is built up by interaction of thousands of genes and proteins.

The start of the Human Genome Project in the late 1980's refocused scientists' attention on new, high-throughput technologies for handling and analysing DNA. Several technologies developed by physicists were evaluated for their potential use in the study of biomolecules. For example, mass spectrometry and scanning tunnelling microscopy are now being developed for analysing and sequencing DNA. Recently, *C. Venter* and colleagues announced at the 10th genome-sequencing conference held in Miami, Florida, in September 1998 that they would have finished the entire human-genome sequencing project by the year 2003. This goal appears now realistic because of the recent development of high-throughput capillary DNA-sequencing machines that can automatically process about 1600 samples per day generating almost a megabase of sequence. Once the genome sequence is completed, it is obvious that highly parallel approaches are needed to assign functions to the endless stretches of G, A, T or C residues.

Classical solid-phase filter-hybridisation techniques like *Southern*, Northern, slot or dot blots are certainly not suited for handling and processing the information contained in the estimated 100000 genes of the human genome (*Fig. 1*). The polymerase chain-reaction (PCR) appears already more efficient because parts of this nucleic-acid amplification method are amenable to automation. The microchip, developed by physicists for computer applications, appears to be the most promising technology to develop tools for high-throughput analysis in biology. The semiconductor industry created the ability to manufacture silicon chips with smaller and smaller features, thus allowing to inte-

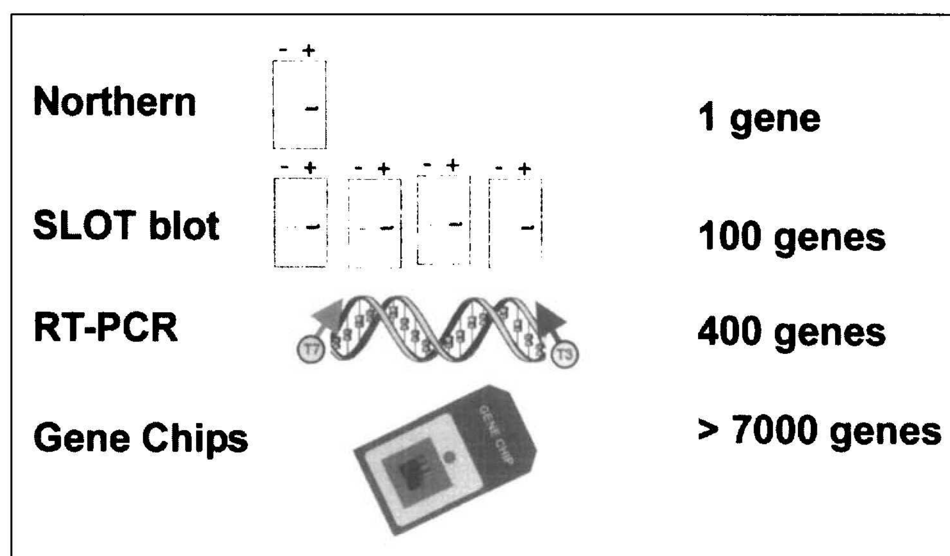

Fig. 1. *Comparison of the throughput capacity of common gene-detection techniques.* **Northern** blots contain total messenger RNA of cells or a tissue size-fractionated and immobilised on a solid support such as nitrocellulose. **Slot blots** are similar but the mRNA is immobilised without previous size-fractionation which results in higher throughput. **RT-PCR** is a variant of the polymerase chain-reaction, which amplifies gene transcripts with specific primer pairs. Because only pipetting and gel electrophoresis are involved, the technique can be partially automated and performed in microtiter plates. **GeneChips** are reversed Northern blots because the probes are immobilised rather than the target RNA. Miniaturisation is the key to the parallel analysis of up to 7000 genes.

*Correspondence: Prof. U. Certa
[a] F. Hoffmann-La-Roche Ltd.
Department PRPG
Grenzacherstr. 127
CH–4070 Basel
Tel.: +41 61 688 53 40
Fax: +41 61 688 14 48
[b] Institute for Medical Microbiology
University of Basel
Petersplatz 10
CH–4010 Basel

CHIMIA 1999, 53, No. 3

grate more and more transistors on a chip of standardised size. This benefit of miniaturisation is now applied to biological sciences by creating multiparallel arrays of DNA fragments on a small chip.

2. Proteomics

Large-scale sequencing projects, DNA chip technologies and rapid genetic-test systems in model organisms all fall under the recently created new branch of science termed 'functional genomics'. Since proteins execute the commands embedded in our genome, high-throughput protein analysis is of equal importance. Mass spectrometry, again a technique developed by physicists, was adapted as an analytical tool in biology and allows for the first time the parallel identification of thousands of proteins separated on a two-dimensional polyacrylamide gel. Tissues, cells or cell fractions are collected and lysed in an appropriate buffer which solubilises proteins. The polypeptides are subsequently resolved on a two-dimensional slab gel using two independent separation principles: charge and molecular weight (see *Fig. 4*). Once separated, individual protein spots of interest are isolated and digested with a protease with known cleavage specificity. The mass of the resulting peptides is then determined by mass spectroscopy using highly specialised instruments designed for high throughput. In parallel, all proteins present in the public databases are 'electronically' digested with the same protease, followed by comparison of the real and the virtual data. A match of two or more peptide masses allows identification of the protein with high confidence. Micro-sequencing can identify proteins not present in the databases and the partial protein-sequence information is usually sufficient to clone the gene. This novel branch of protein analytics, termed proteomics, allows direct visualisation of up- or downregulated protein expression in two or more samples. Typically, a normal and a cancer cell are compared (see below) or situations in which one sample received a drug and the other did not. Proteomics also detects efficiently post-transcriptional modifications of polypeptides such as phosphorylation or glycosylation.

Decoding of our genome will, thus, require both genomics and proteomics, but also functional genetics which deals with, *e.g.*, mouse knockout mutants. Given the complexity of our genome with an estimated number of 100 000 genes, both the detection levels and the throughput capacity of our current techniques will have to improve in order to understand the molecular functions stored encrypted in eukaryotic genomes.

3. Gene Chips

The current development of gene chips has striking similarities to developments in the computer industry that occurred over the last couple of decades. In the 80's, big reels with magnetic tapes were used to store data. Computers were big, slow and used up the space of an entire laboratory or office. In addition, these rooms were equipped with air conditioning in order to avoid overheating of these monster computers. In 1985, the appearance of personal desktop computers revolutionised the entire industry. These machines stored data on small hard or floppy disks and were comparatively easy to use. Hard disks in those days had a typical storage capacity of about 20 megabytes, which is today's average file size of a single bitmap image, for example. At the same time, the DNA and protein databases started to develop. The very first version of the entire EMBL database contained about 800 DNA sequences. Twenty 5.25-inch floppy disks were sufficient for data storage and it took overnight to screen these sequences. From then on, a rapid development occurred. The processors became not only more powerful, but also faster and, more importantly, much smaller. In the 1990' small portable computers appeared and today these instruments fit into the pockets of our jackets. Soon, entire computers will fit on a silicon chip. Thousands of these 'single-chip computers' can be incorporated into a single instrument, and the power of such a computer is beyond any imagination. This example illustrates that miniaturisation is in fact the key to highly-parallel, complex, high-throughput technologies such as gene chips.

Many different DNA-chip techniques have been developed within the last few years, and we selected two for a more detailed presentation because they are applied in our laboratories. We will first discuss PCA (plasmid chip-array) gene chips developed at *Roche* and oligonucleotide arrays, the so called *GeneChips*™ invented and marketed by the California-based company *Affymetrix*.

4. Plasmid Chip-Arrays (PCA)

It has been demonstrated, some years ago, that denatured, double-stranded DNA, spotted onto charged glass surfaces with microjets, could efficiently hybridise with labelled probes [1]. Since the samples are usually labelled with fluorescent dyes, glass is the ideal matrix because disturbing autofluorescence is low. We decided to spot plasmids, rather than DNA fragments generated by the polymerase chain-reaction (PCR) or synthetic DNA, because plasmid isolation from bacterial cell cultures can be automated. Briefly, a robot picks colonies containing recombinant plasmids from an agar plate and inoculates liquid cultures. A second robot isolates the plasmid DNA from these cultures using affinity chromatography. After denaturation, the samples are microspotted using a piezojet. This device deposits 200 picoliters of aqueous plasmid-DNA solution on the charged glass surface independent of the viscosity of the sample. It is noteworthy that a single microliter of solution is sufficient to produce 5000 chips. In order to control the absolute amount of DNA and chip quality, each PCA consists of two identical fields with up to 1600 recombinant plasmids, including appropriate experimental controls (*Fig. 2*). The left array on the chip is hybridised with a fluorescent plasmid-backbone probe and the right one with the sample of interest, usually messenger RNA. The plasmid signal is proportional to the total amount of plasmid DNA deposited on the chip, whereas the spot intensity on the array hybridised with the sample reflects the expression level or abundance of a gene. Although the spots have a diameter of only 150 µm, they still contain about 10^7 molecules of the recombinant plasmid DNA. The target molecules on the chip are, thus, present in excess with respect to the copy number in the sample, which is important for the dynamic range, signal quantification and the final interpretation of the RNA-expression data.

The PCA technology is suited for either transcript imaging or gene mapping because any kind of recombinant plasmid can be spotted and all types of nucleic-acid probes can be used. Furthermore, the whole process is completed within one week from colony picking to data analysis without any external resources or dependencies. Compared to PCR-based approaches, PCA does not require any knowledge about the sequence of the genes to be analysed, and the plasmid samples can come from clone collections, cDNA, or genomic libraries. For example, entire libraries can be screened differentially in micro format using standard protocols, rather than using large nylon filters and radioactive probes. In addition, one can

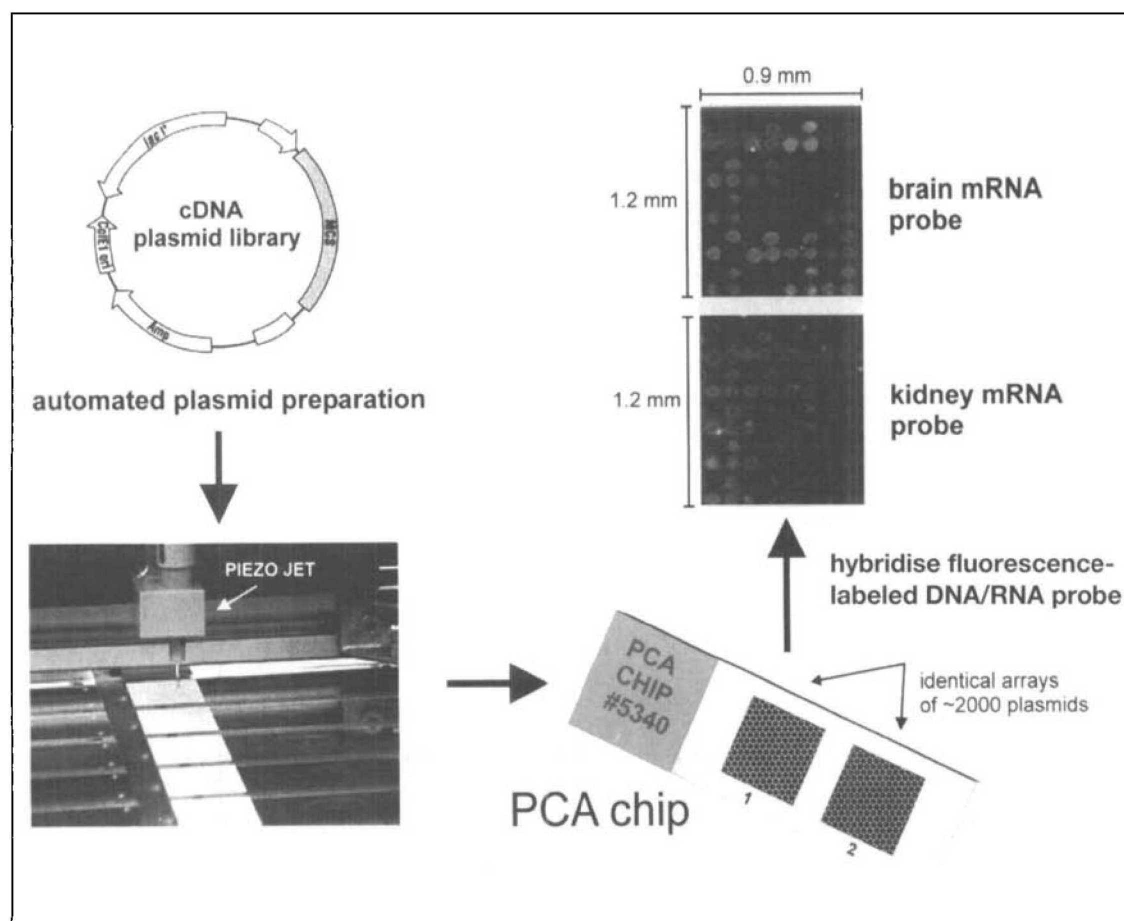

Fig. 2. *Outline of the PCA technology* (see text for details)

replace fluorescent labels with normal colour substrates used for *in situ* hybridisation or non-radioactive detection of nucleic acids. This eliminates the need for expensive laser scanners, and the chip analysis and data collection can be performed using a standard light microscope connected to a video camera and a computer. PCA may not be as precise and sophisticated as other chip technologies, but the technology is flexible, relatively cheap and, most importantly, it does not require any sequence information about the genes to be analysed.

5. Oligonucleotide *GeneChips*™

The *GeneChip*™ technology is different from PCA arrays and was invented by the California-based biotech company *Affymetrix* [2–4]. The technology relies on oligonucleotide, rather than plasmid probes, which implies that these technologies complement, rather than compete with each other.

5.1. *Chip Production*

The roots of the technology go back to light-directed computer chip production. Using masks, several thousand oligonucleotides are built up in parallel on a glass chip of only 1 cm^2. If a light source is directed to a small 50 μm^2 area on the chip through a mask, a photolabile protection group is removed and the first nucleotide of an oligonucleotide probe is coupled at that particular position. The chip is now submerged in a reaction solution containing the desired 3'-protected nucleotide residue. Using another mask, other areas of the chip are deprotected, and, thus, several repetitions of this process allow the build-up of up to 200000 different oligonucleotides in parallel. During the production process, 80 chips are made in parallel on two glass plates or wafers, which means that all chips of such a wafer are identical due to the production process. This also implies that the sample- and not the chip- introduces variation in an experiment.

5.2. *How Gene Chips Work*

On a single chip, 64000 22-mer oligonucleotides are synthesised in parallel which is sufficient to analyse 1500 genes. Let's say, the coding region of gene X has a length of 1000 bases. From this sequence, twenty unique stretches of 20 nucleotides are selected (*Fig. 3*). These have to meet several criteria, for example sequence uniqueness, or lack of potential secondary-structure formation. To control the specificity of hybridisation, the same 'set' of oligonucleotides is synthesised on the chip, but with a mismatch nucleotide in the centre of the sequence. RNA is extracted, amplified and labelled by *in vitro* transcription and applied to the chip. Only expressed genes will find their probes and hybridise more strongly to the perfect-match probes than to the mismatch probes. For expression analysis, the intensity of the sum of mismatch signals is subtracted from the match signal. This average difference is comparable with the band intensity on a conventional blot, but based on multiple measurements. If necessary, this analysis method also allows the electronic modification of the hybridisation conditions. Multiple measurements and the possibility of fine-tuning of the analysis are unique features of the *Affymetrix* technology. Most other microarray techniques, including PCA, rely on single hybridisation to long, unique sequence stretches.

6. Chip Application: Protease Expression and Apoptosis in Tumour Cells

Apoptosis is defined as a process of programmed cell-death occurring in all higher organisms. Currently, the mechanisms of apoptosis are under extensive investigation. The inability of a cell to execute this suicide program, for example

after failing to repair DNA damage, is causatively linked to cancer. The so-called oncogenes, originally discovered in RNA tumour viruses, can transform normal cells into cancer cells. Transformation of a normal cell into a tumour cell can also be achieved *in vitro* by transfecting a recombinant plasmid carrying an oncogene. Using gene-chip and proteomic technologies, the normal precursor cell (PB-3c) and the transformed malignant tumour cell (V2D1) transfected with the H-ras oncogene can be compared at the molecular level in order to discover pathways or events, which may explain the transformation process. For example, the normal PB-3c cell requires the cytokine interleukin-3, which is produced after oncogene transfection in an autocrine fashion, a typical property of tumour cells. Using the more conventional differential-PCR display technology, we have already identified a small subset of genes differentially expressed in the two cell types [6]. We, thus, grew up both lines, prepared protein and RNA samples, and subjected those either to protein analysis on 2D-gels coupled with mass spectroscopy, or gene-chip analysis using the *Affymetrix* system.

6.1. *Proteome Analysis*

Fig. 4 (top) shows the relevant part of the 2D-gel stained with colloid *Coomassie* blue. According to that analysis, BTF3a (a transcription factor) and ICE3 (a caspase involved in apoptosis) are upregulated, whereas HMG2 (a chromosomal protein) are downregulated in the cancer cells and the level of several other proteins displayed on this part of the gel remains unchanged. All spots including ICE3 were identified by mass spectrometry as described above. The active, 27 kDa form of ICE3 is usually generated by proteolytic cleavage of a 30 kDa precursor protein. The protein analysis reveals, that only the inactive precursor of ICE3 is detected and accumulates in the cancer cells.

6.2. GeneChip™ *Analysis*

Fig. 4 (bottom) shows the scanned chip images of the two samples. Just by visual inspection, the two samples appear virtually identical with one exception (high-lighted by the white box). This gene encodes the protease granzyme B, which is turned off as a consequence of ras-transformation. As it turns out, granzyme B activates, under normal conditions, ICE3 by proteolytic cleavage [5]. The combina-

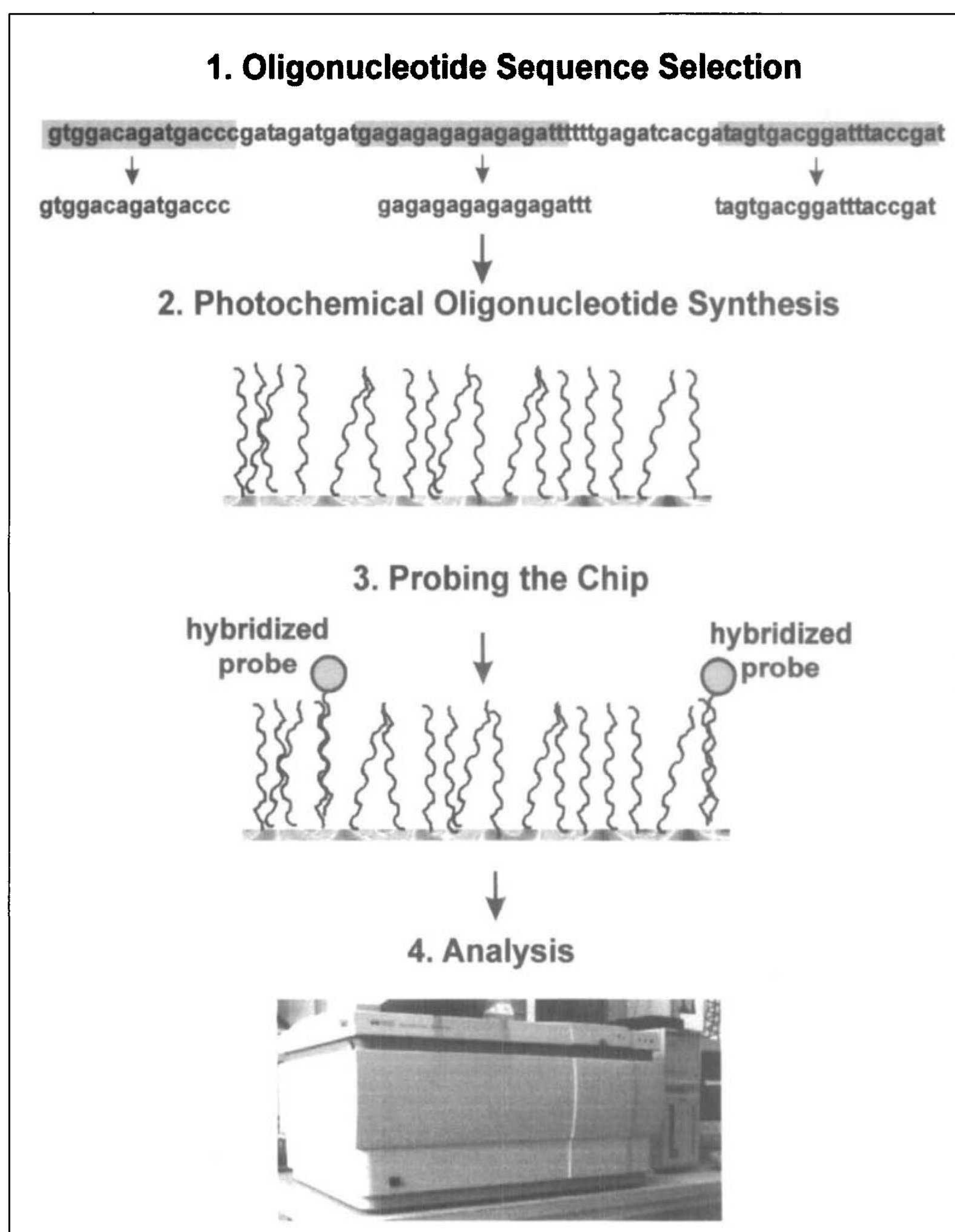

Fig. 3. *Outline of the* GeneChip™ *technology.* From the sequence of the gene to be analysed appropriate sequences are selected as probes (1). These sequences are now synthesised chemically on a small glass chip in form of oligonucleotides (2). This microarray of 64000 oligonucleotides/cm^2 is probed with a fluorescent sample (3) and molecules that have successfully hybridised are detected by laser scanning (4).

CHIMIA 1999, 53, No. 3

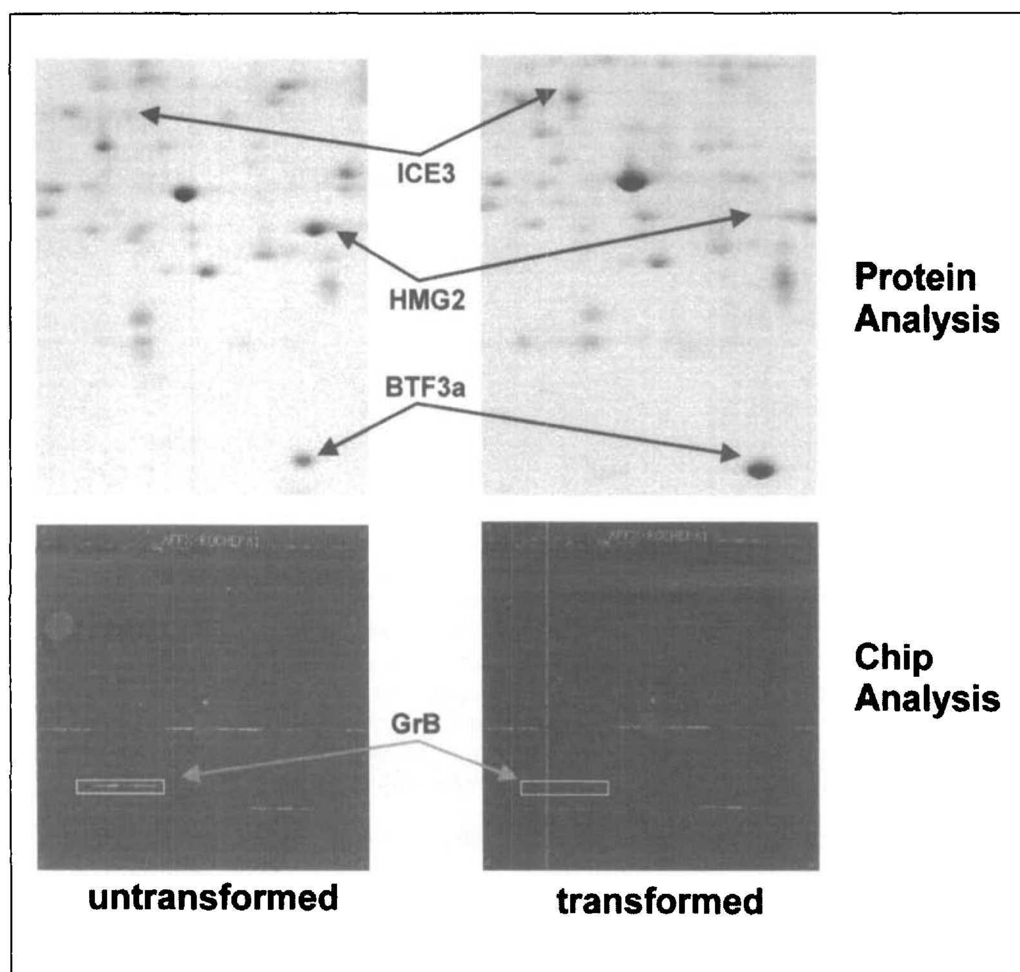

Fig. 4. *Application of proteomics and* GeneChip^{TM} *technologies for the investigation of apoptosis.* Proteins or RNA coming either from normal mouse mast-cells or ras-oncogene-transformed cells were analysed either on 2D-gels (*top*) or on *Affymetrix GeneChips*^{TM} (*bottom*). Relevant spots or genes are labelled with arrows; see text for details.

tion of proteomics and genomics confirms this mechanism: shut-down of granzyme-B expression results in an accumulation of the inactive ICE3 precursor and, as a result, the caspase is inactive. The protease activity is likely to be important for the apoptotic pathway and its lack results in malignant, uncontrolled growth of the cells.

7. Outlook and Future Gene-Chip Applications

DNA-chip technologies will certainly speed up functional genomics and we anticipate that this technology represents one of the major break-throughs in the biology of the future. The challenge is to detect the functional relationships of the 100000 genes encoded by our genome, in order to understand how a multicellular organism is built-up. How difficult this scientific endeavour is going to be emerges already from research in bacteria where several complete genomes and consequently all proteins are known. It is still unclear how they interact although the first, novel pathways are about to be elucidated. A sound knowledge of the interactionpathways is important for understanding the molecular mechanisms responsible for drug resistance or pathogenicity. This illustrates that gene chips or proteomics alone will not answer all of our questions. Whether pro- or eukaryotes; we need additional tools such as genetics, biochemistry or cellular assays to build up the complete picture.

Another area in which gene chips will become indispensable tools is the field of disease diagnostics and management. It is already possible to diagnose genetic risk-factors for cancer, or to determine sequences of HIV viruses with high reliability using *Affymetrix* chips. Since the capacity of DNA chips is high, multiple risk factors or mutations for genetic disorders can be analysed with a single blood sample. Physicians can then apply preventive medicines before the patient becomes sick. *Alzheimer*'s disease and certain cancers are examples where such genetic risk-factors are involved. Gene chips are already indispensable research tools and the first diagnostic applications start to emerge. Yet, the technology is still quite complex and also expensive. Nevertheless, this will probably change based on experiences from the past: the polymerase chain-reaction (PCR) was discovered more than ten years ago. Scientists were sitting in front of three waterbaths set at different temperatures; the tubes with the reaction mixture were in a rack and shuttled between the baths according to the protocol. A single experiment took at least four hours of patience and concentration until the amplified DNA sample was ready for gel analysis. Today, a hand-held thermocycler is injected with a blood sample and about 45 min later the result '400 virus particles/ml' is displayed on a LCD screen. Who had predicted such a development ten years ago?

Received: November 2, 1998

[1] M. Shena, D. Shalon, R.W. Davis, P.O. Brown, *Science* **1995**, *270*, 467.
[2] S.P.A. Fodor, R.P. Rava, X.C. Huang, A.C. Pease, C.P. Holmes, C.L. Adams, *Nature (London)* **1993**, *364*, 555.
[3] L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, *Nature Biotechnology* **1997**, *15*, 1359.
[4] A. de Saizieu, U. Certa, J. Warrington, C. Gray, W. Keck , J. Mous, *Nature Biotechnology* **1998**, *16*, 45.
[5] S. Wang, M. Miura, Y. Jung, H. Zhu, V. Gagliardini, L. Shi, A.H. Greenberg, J. Yuan, *J. Biol. Chem.* **1996**, 271, 20580.
[6] M. Buess, O. Engler, H.H. Hirsch , C. Moroni, *Oncogene* **1999**, in press.