CHIMIA 2001, 55, No. 1/2

Chimia 55 (2001) 35–39

ISSN 0009–4293

DNA and Protein Microarrays and their Contributions to Proteomics and Genomics

Markus Ehrat* and Gerhard M. Kresbach

Abstract: Knowledge in genomics and proteomics has exploded in the last two decades. This is in part due to key developments that have revolutionized the possibilities of bioanalytics such as the introduction of polymerase chain reaction (PCR) in the mid 80s that formed the base for the massively parallel sequencing of the genomes.

A few years ago DNA and protein microarray analysis were added to the toolbox of life sciences analytics. These technologies already proved to be ideal tools for the identification of gene targets, the simultaneous measurement of the expression of a high number of genes or proteins, and the increase of the level of understanding of the biological functions of genes and proteins. A small number of experiments are now sufficient to obtain information on gene or protein expression which could not be obtained by using conventional bioanalytical technologies or which required an extremely high experimental effort. In the future applications, high sensitivity DNA and protein microarrays will allow low abundant genes and proteins to be monitored that so far have been inaccessible to current microarray technologies and thus will generate a new dimension of genomic and proteomic information.

Keywords: Bioanalytics · DNA microarray · Gene expression · Genomics · Planar waveguide · Protein expression · Protein microarray · Proteomics

Introduction

It is generally accepted that the information stored in about 25 000–50 000 genes is required to build and to operate such a complex biological system as a human being. Taking into consideration possible translational variants, different post-translational steps of protein synthesis as well as the transformation in metabolic processes, as many as 20 million different peptides and proteins are expected to exist of which the actual biological functions are to a large extent not yet fully known. DNA and protein microarrays that allow the efficient and highly parallel monitoring of a vast number of biological interactions will play a key role elucidating the function of such a complex network (Fig. 1).

*Correspondence: Dr. M. Ehrat
Zeptosens AG
Benkenstr. 254
CH-4018 Witterswil
Tel.: +41 61 726 81 81
Fax: +41 61 726 81 71
E-Mail: markus.ehrat@zeptosens.com
www.zeptosens.com

Regardless of the fact that the determination of the sequence of the human genome is almost complete, information derived from DNA/RNA *and* from protein analysis is required in order to understand the functions of such a complex and highly regulated biological system. Once the function is understood, a large number of different biological marker molecules have to be measured in parallel for the monitoring the biological system.

To generate biologically relevant information often only a very limited amount of material is available: *e.g.* in single cell analysis of tissue where material is obtained by laser micro-dissection to generate locus specific information on gene expression levels. Miniaturization of analytical technologies down to the micrometer and picoliter scale is mandatory to handle minute amounts of samples and reagents basically free-of-loss in a fast and sensitive manner.

The technological revolution in genome research started in 1985/1986 with the introduction of the polymerase chain reaction by Ehrlich and coworkers [1]. For the first time, genetic material could be amplified on a large scale thus making massive sequencing efforts possible.

Many breakthroughs in genomics and proteomics in the last years are based on the highly parallel generation and analysis of biological information. This has been made possible to a large extent by only a few key developments in analytics and bioanalytics: miniaturization of the devices, parallelization of the measuring procedure, utilization of biospecific interaction mechanisms and – as a consequence thereof – the development and use of DNA and protein microarrays.

DNA Microarrays

Due the complexity of an organism and the processes occurring there and the stringent requirements with respect to drug safety and efficacy, more and more tasks in pharmaceutical research and development require the monitoring of a large number of biological marker molecules to find and correctly describe drug mechanisms or to project the possible action of novel drug candidates.

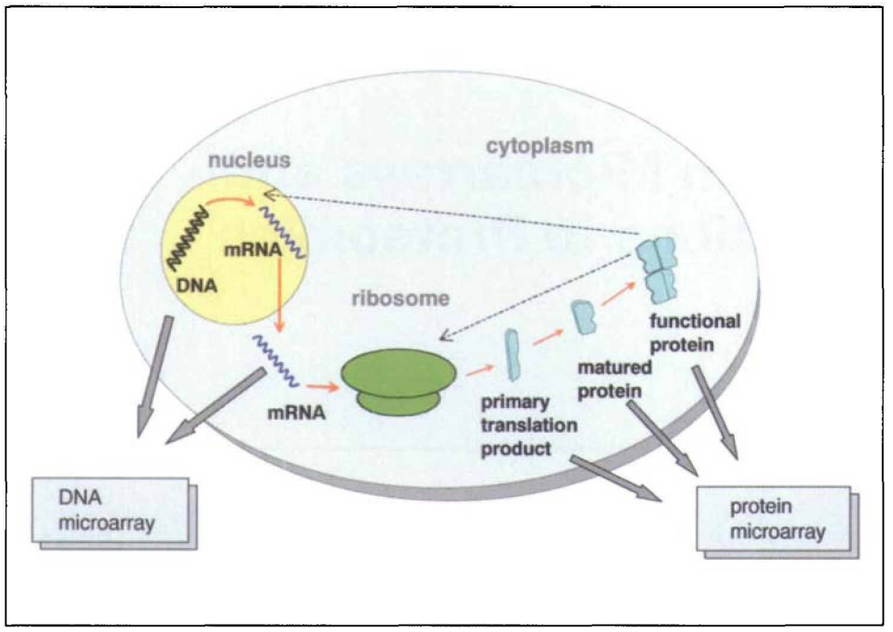

Fig. 1. Intracellular gene and protein expression mechanisms. DNA, mRNA as well as tRNA and rRNA (the latter two not shown in the diagram) can be analyzed and monitored by using DNA microarrays, the proteins in their different maturation states by using protein microarrays.

Today's most important DNA microarray applications are gene screening for the detection of novel drug targets, determination of gene mutations and single nucleotide polymorphism and gene expression monitoring for the identification of disease specific genes and for the determination of toxic effects of drug candidates. Many future microarray applications might be in the field of human diagnostics, *e.g.* monitoring the genetic patterns of individuals as a base for individualized drug therapy, diagnosis of the genetic predisposition to develop certain diseases or the fast and sensitive identification of bacterial and viral pathogens. For some of these tests, however, considerable ethical concerns do exist. The future will show if such tests will be widely accepted.

DNA microarrays typically consist of a flat glass substrate onto which an array of cDNA or oligonucleotides are deposited as spots of a diameter between 10 and 200 µm. To identify specific mRNA, or to determine their copy numbers, such RNA is extracted from the cell and reverse transcribed into cDNA. During this copying process, fluorescently labeled nucleobases are introduced which can have a different color for sample and reference experiments. The labeled cDNA is brought in contact with the array surface and hybridizes to the recognition elements. Specificity can be achieved through careful selection of hybridization temperature, buffer composition and washing conditions. The amount of bound DNA carrying a fluorescent label is typically determined with a reader system scanning the individual spots.

For the production of *high-density microarrays* photolithography methods are applied. Microarrays featuring several tens of thousands different on-surface synthesized, short oligonucleotide recognition elements can be obtained with a feature size of about 20 µm. They allow the parallel detection of several thousand genes in a fast and relatively simple fashion. Typically about 20 different feature elements are required to characterize *e.g.* a single-base mutation within a segment of a gene.

For the production of *mid-density microarrays* non-contact deposition methods, comparable to ink-jet printing or contact deposition methods, using pins with a diameter of about 100 µm, are used. Pre-synthesized oligonucleotides or cDNAs amplified from genetic material are used as recognition elements – one to three recognition elements are typically required to determine the expression level of a gene.

Both types of microarrays have been successfully applied in different fields of genomic research. They reach their limits, however, when genes of low abundance have to be monitored or the available sample material is very limited like material derived from laser microdissection, miniaturized cell cultivations, or needle biopsies.

Target and signal amplification schemes often are accompanied with a significant bias in the generated information: *e.g.*:

- Different genes or gene fragments are PCR amplified at a different scale – the original expression information is distorted
- Amplification occurs only if a certain number of molecules is present – low abundant information is suppressed
- Contaminating molecules are amplified – a significant noise level is introduced
- Amplification reactions are not robust – results vary within a set of experiments
- Label incorporation is not reproducible at low concentrations – the signal is influenced by the composition of the nucleobases.

As indicated in Fig. 2, a cell typically contains about 10 pg of total RNA, of which about 100–500 fg is mRNA. 0.1% of this mRNA consists of high abundant genes with more than 5000 copies per cell which are accessible to currently available microarray detection systems. However, 95% of the mRNA, are composed of low abundant genes with less than 10 copies per cell which can be monitored only with high effort in target and signal amplification. The consequence of the inability to correctly measure low abundant genes is that up to 95% of the information encoded in a cellular gene expression pattern is lost or not interpreted in a correct manner.

To overcome the current limitations of microarray detection with respect to signal intensity, signal-to-noise ratio and linearity of the signal, Zeptosens is developing a planar waveguide (PWG) technology (Fig. 3) as a high sensitivity measurement platform. The particular advantage of PWG is the surface-confined, extremely strong electromagnetic field which allows the efficient and selective excitation of fluorescence labels on

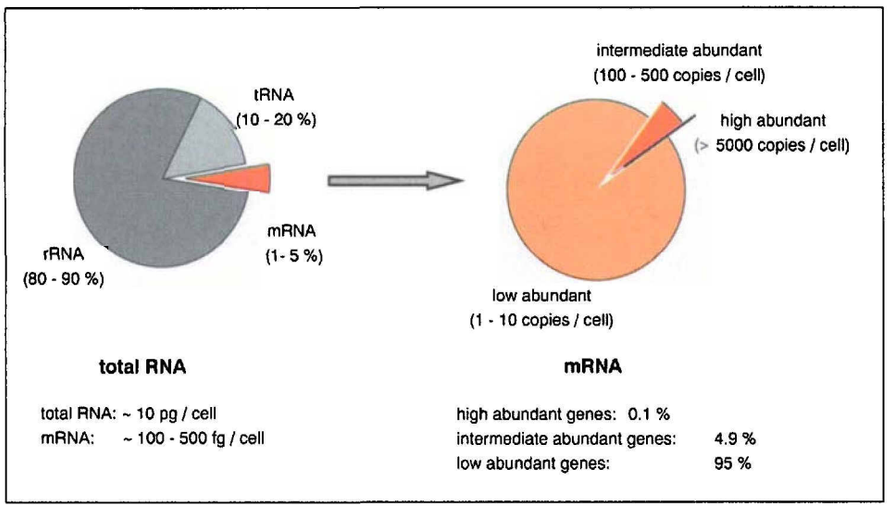

Fig. 2. Typical distribution of type and abundance of RNAs. Total RNA consists only to a small percentage of mRNA; high abundant genes that are typically monitored using today's commercially available DNA microarrays and microarray read-out systems reflect only to a very small fraction of the information encoded in the genome.

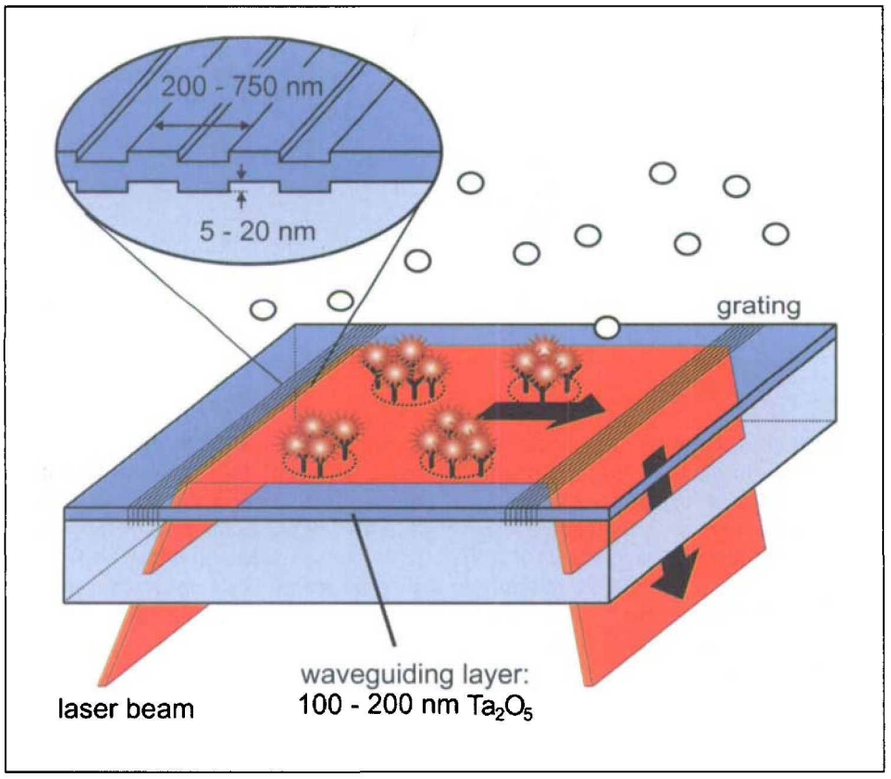

Fig. 3. Planar waveguide (PWG) detection – the principle behind the Zepto™READER.

the surface: Laser light is coupled *via* a diffractive grating into a thin film of high-refractive-index material deposited on a transparent support. An evanescent field is created, which has a limited penetration depth of about 200 nm into the adjacent medium. Thus, only surface-confined fluorophores are excited and do emit fluorescence signal whereas fluorophores in the bulk medium are not excited. As a consequence, the background signal derived from chemical noise can be effectively reduced. These properties result in a significant increase in the signal/noise ratio compared to conventional optical detection methods and permit highly selective and sensitive measurements. Monitoring can be performed easily in solutions which further contributes to the precision of the measurement.

Currently, compared to the sensitivity of conventional confocal epifluorescence scanners, an improvement factor of up to 100 can be achieved by using the Zepto™READER. Using this technology, abundant genes expressed in as little as 100 cells could be monitored. Instead of using *e.g.* 10 to 100 mg of brain tissue (equal to 10–100 µg of total RNA) for the analysis of rare messages [2] less than 25 µg (equivalent to less than 25 ng of total RNA or less than 1 ng mRNA) are required for the monitoring of gene expression (Fig. 4). This allows sampling *via* needle biopsies, laser dissection, and cultivation in miniaturized systems.

Protein Microarrays

Compared to gene expression, the monitoring of proteins is significantly more complex but is closer to the real nature of disease and drug action. In contrast to DNA microarrays where the nature of the potential biorecognition partners is limited to polynucleotide sequences or chemically similar polymers protein microarrays can be designed using quite a number of chemically different ligand–receptor pairs such as antibody–antigen, enzyme–substrate, cell membrane receptor–ligand, peptide–protein or protein–DNA/RNA. Hurdles for the development of protein microarrays are based in the specific properties of the recognition elements *e.g.* an antibody might not be able to distinguish the molecular sub-variants, an antibody will exhibit a cross reactivity towards other proteins of significantly different function, the binding conditions might vary significantly among the different recognition pairs or the tertiary structure of proteins might be altered or lost during immobilization on surfaces.

There are already many fields in pharmaceutical research and development where sets of biological marker molecules are used to monitor health status,

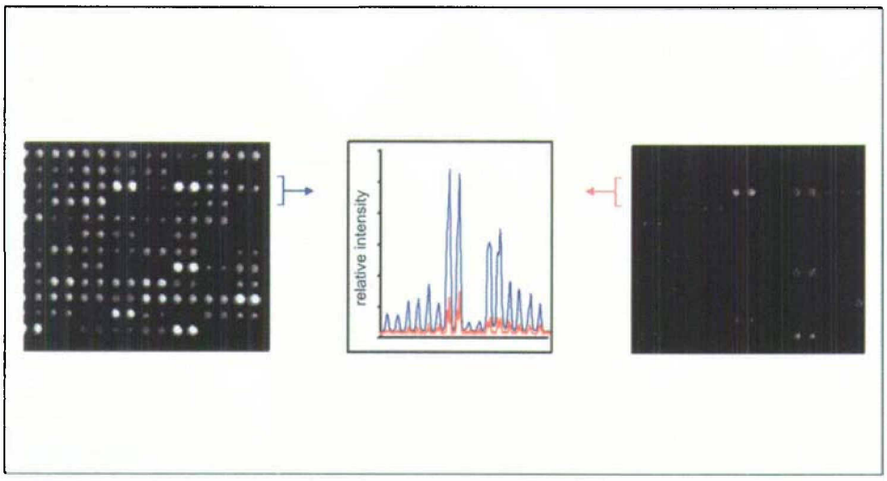

Fig. 4. Comparison of microarray readout results obtained by using a PWG based Zepto ™ READER (left) and a state-of-the-art confocal fluorescence scanner (right). The signal-to-noise ratio is about 60 fold increased when using the Zepto™READER (center, upper trace) compared to the epifluorescence scanner (center, lower trace). mRNA derived from 25 µg of mouse brain tissue was monitored in this experiment.

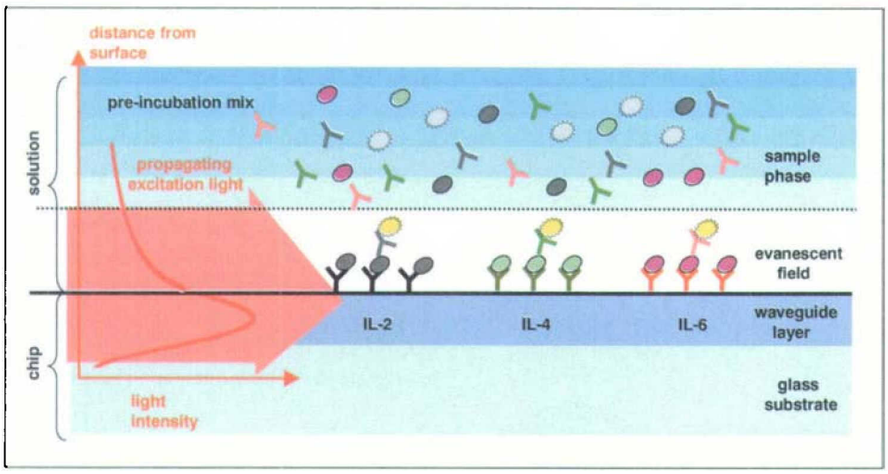

Fig. 5. Schematic visualization of the triplexed assay used for the detection of the three human Interleukins hIL-2, hIL-4, and hIL-6.

disease progress or success of a treatment. As an example in inflammation, a set of different cytokines already play a significant role as biological markers to determine the status of an infection. Up to now, however, the biological markers are monitored in ELISA type assays on a one by one basis. This approach requires relatively high amounts of sample material, time-consuming and tedious sample preparation and assay performance procedures, consumes large amounts of costly reagents like antibodies and labels, and, last but not least, might result in contradicting conclusions.

Currently, the parallel detection of expressed proteins is limited in applicability and throughput and requires significant experimental efforts. 2D-gel electrophoresis with subsequent analysis of individual segments has been performed using mass spectrometry. ELISA assays are performed in a 96-well format allowing the parallel determination of about 30 samples. Radioactive labels and chemiluminescence processes were developed but are either critical in terms of ecology or handling. Time-of-flight mass spectrometry was introduced as a sensitive analyzer but exhibits drawbacks with respect to throughput and bias towards lower molecular weight compounds.

It is only very recently that the microarray format already established for DNA analysis has been introduced into protein analysis. Protein microarrays will be able to measure protein–ligand interactions in parallel and in a miniaturized format, eliminating most of the above mentioned drawbacks. As soon as proteins of natural origin are investigated as target molecules, the sensitivity of the detection approach is even more critical as – in contrast to polynucleotides – no simple and efficient target amplification schemes are available.

Multiplexed protein assays can be subdivided – comparable to DNA analysis – into two major fields of applications:

1. Analysis of protein expression patterns of a large number of known and unknown target molecules
2. Analysis of function, activity, and concentration of a subset of target molecules

For the first approach MacBeath and Schreiber [3] have shown that – using printing technologies comparable to these utilized to manufacture DNA microarrays–protein microarrays featuring several thousands of different proteins can be readily produced and – using specific antibodies – individual proteins and protein families can be successfully addressed and identified.

Zeptosens is developing a robust multiplexed assay format utilizing the specific advantages of PWG detection (Fig. 5). Antibodies are spotted onto the chemically activated planar waveguide layer that forms the bottom of a reaction and detection compartment. Sample liquid containing antigens, labels, and tracer antibodies are pre-reacted and given into the sample compartment. As soon as the bioaffinity reaction takes place, monitoring of the bonding process is possible without any additional washing-steps what makes this approach suitable for kinetic measurements.

Fig. 6 shows an example where three different human interleukin antibodies are monitored in parallel with individual sensitivities down to the pg/ml range.

Compared to conventional ELISAs, a multitude of advantages can be obtained

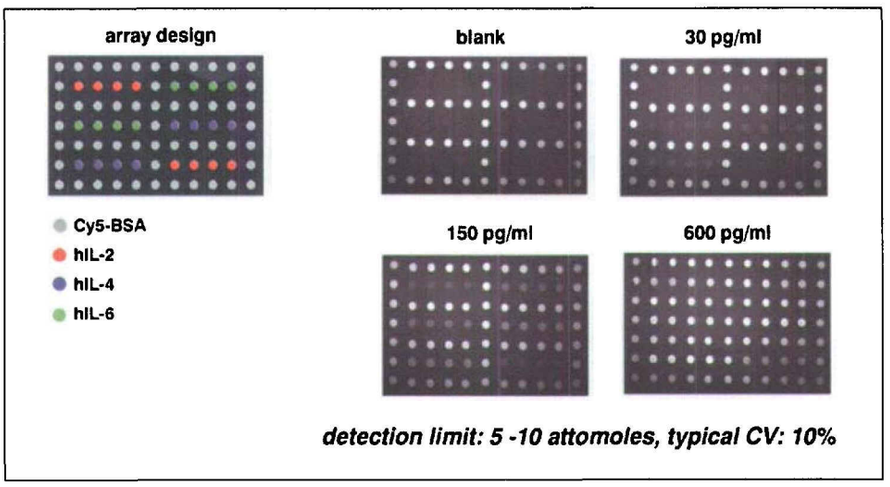

Fig. 6. Array design and experimental outcome of a triplexed human Interleukin assay. Linear calibration curves down to the low pg/ml range could be obtained in these experiments, limits of detection typically were found to be 1 pg/ml.

Table.

ELISA	ZeptoMARK
• single biological marker per well	• up to 10 biological markers per well
• replications in multiple wells	• several replications in one well
• 150 µl sample volume	• 15 µl sample volume
• 6000 µm 'spot diameter'	• 200 µm spot diameter
• LOD = 1–10 pg/ml, 6–60 attomole detectable	• LOD = 1–10 pg/ml, 600 zeptomole detectable

ranging from substantial reduction in sample and reagent consumption due to improved sensitivity of the readout to a significant improvement in the quality of the biological and analytical data obtained because several biological markers can be determined in multiples in the same microenvironment in parallel (Table).

The degree of multiplexity is primarily limited by the selectivity of the recognition element. When sacrificing selectivity to a certain extent high sensitivity protein microarrays with densities of several hundreds of features per cm^2 can be readily produced.

Conclusions

DNA microarray technology has experienced a tremendous boost in the last five years and is already an accepted technology in the research and development laboratories of all major pharmaceutical and agricultural companies. Protein microarrays are being developed in several academic and industrial laboratories and will evolve as one of the major platforms for proteomic research. Especially the high sensitivity DNA and protein microarrays will have a significant impact on genomic and proteomic activities as they allow:

- Sensitive analysis with no or only linear signal / target amplification
- Investigation and monitoring of low abundant genes and proteins
- Tissue specific investigations of protein and gene functions

DNA and protein microarrays will create an avalanche of new data. Highly sophisticated bioinformatic approaches will be required to convert these data into useful information. This information will further contribute to the elucidation of the biological functions of genes and proteins and thus open the door to novel approaches in drug development, patient treatment and therapy. Thus DNA as well as protein microarrays will help to improve the quality of our lives.

Received: January 16, 2001

[1] S.J. Scharf, G.T. Horn, H.A. Ehrlich, *Science* **1986**, *233*, 1076–1078.

[2] D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, *Nature Genetics Suppl.* **1999**, *21*, 10–14.

[3] G. MacBeath, S.L. Schreiber, *Science* **2000**, *289*, 1760–1763.