

MINIATURIZATION IN FUNCTIONAL GENOMICS AND PROTEOMICS

Sascha Sauer, Bodo M. H. Lange, Johan Gobom, Lajos Nyarsik, Harald Seitz and Hans Lehrach

Abstract | Proteins are the key components of the cellular machinery responsible for processing changes that are ordered by genomic information. Analysis of most human proteins and nucleic acids is important in order to decode the complex networks that are likely to underlie many common diseases. Significant improvements in current technology are also required to dissect the regulatory processes in high-throughput and with low cost. Miniaturization of biological assays is an important prerequisite to achieve these goals in the near future.

SYSTEMS BIOLOGY

The study of the complex interactions that occur at all levels of biological information — from whole-genome sequence interactions to developmental and biochemical networks — and their functional relationship to organism-level phenotypes.

Max Planck Institute
for Molecular Genetics,
Department of Vertebrate
Genomics, Ihnestrasse 73,
D-14195 Berlin, Germany.
Correspondence to S.S.
e-mail: sauer@molgen.
mpg.de
doi:10.1038/nrg1618

The information encoded in the genome comprises genes, the protein products of which mediate most of the functions in organisms, and control elements¹. Proteins were thought to be the most important effectors in the cells, although recently non-coding RNAs have also been identified as important players in regulatory processes². Proteins and RNAs rarely function alone, but rather in complexes or in networks. In addition, genomic control elements contribute to gene-regulatory networks.

The regulation of biological processes depends on multiple parameters of a pathway. Many complex diseases can be regarded as perturbations at one of many points on a pathway that controls a normal, functionally intact biological process³. Therefore, real progress in diagnosis and therapy of many common diseases will require the identification and detailed analysis of many and possibly all components of these complex biological processes.

Large data sets, comprising functional genomics, proteomics and metabolomics data, as well as biochemical information, provide the basis for SYSTEMS BIOLOGY⁴. Most of these data are still being generated *in vitro*, although *in vivo* data will have an increasingly important role in the future. Integration and handling of these many data types through modelling will be the only way to convert our current knowledge into a predictive capability for improving the therapy of many complex diseases^{1,4}.

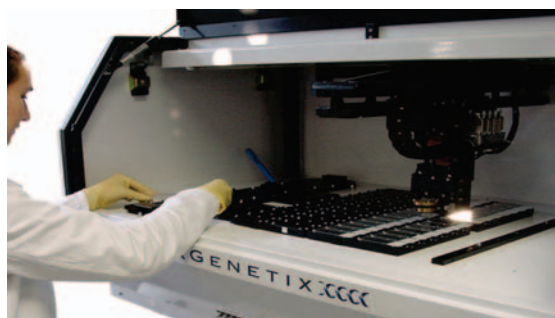
Cost-effective generation of data using large numbers of samples, analysis of limited amounts of biological material and handling of intrinsic biological heterogeneity will be necessary for this systems biology approach, and can be achieved through parallelization and miniaturization of biological assays. The parallelization of oligonucleotide hybridization assays was a key step in the development of genome-wide technologies^{5–8} (FIG. 1). Recently, protein array formats have been used and will certainly improve the parallel analysis of proteins^{9,10}. Miniaturization can complement these developments by reducing the volume and increasing the density of assays. It can also help when analysing larger numbers of samples in parallel. The scale of natural substrates that can be analysed (BOX 1) and various molecular biological devices are summarized in TABLE 1. The ultimate goal of miniaturization involves analysis of smaller, more homogeneous sections of complex samples, reduction of costs and preservation of precious samples.

In this article, we will mainly discuss a selection of miniaturized *in vitro* and *in situ* procedures for the parallel, sensitive and discrete analysis of nucleic acids and proteins.

Reducing complexity

Techniques such as cell sorting¹¹, manual dissection¹² and LASER CAPTURE MICRODISSECTION (LCM)¹³ have been used to isolate homogeneous cell samples, an

0.25–1 nl of DNA solution is printed on a slide, creating spots that range from 100 to 150 μm in diameter



A typical fluorescence read-out of highly parallel hybridization on a cDNA microarray

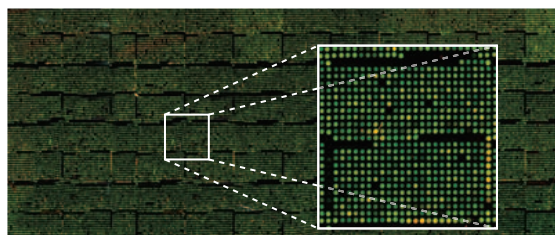


Figure 1 | DNA microrarrays. DNA microarrays that are based on non-porous solid supports such as glass have paved the way for highly parallel miniaturized analysis of biomolecules. Using robotic workstations, ~0.25–1 nl of DNA solution are printed on a slide, creating spots that range from 100 to 150 μm in diameter. Successive samples are then spaced to avoid contact between adjacent spots, with approximately 200 μm between each spot⁷. Currently, about 50,000 cDNAs can be robotically spotted onto a microscope slide and hybridized with a fluorescently labelled probe. Using photolithographical masking techniques, arrays that have 400,000 distinct oligonucleotides have been produced, each in its own 20- μm^2 region (today, this can be as small as ~5 μm^2) (REF. 8). After spotting, nucleic-acid samples can be hybridized on these pre-structured microarrays. Hybridization events are monitored using fluorescence scanners. A typical fluorescence read-out of highly parallel hybridization on a cDNA microarray is shown in the bottom panel.

important prerequisite in many biological studies. However, even cells of the same type usually represent different stages of physiological processes. The complexity of biological material is currently also reduced by applying labour-intensive biochemical CELL FRACTIONATION and subcellular isolation steps before carrying out the analysis of a set of molecules or single molecules¹⁴.

Amplification techniques such as PCR have succeeded in reducing the amount of sample needed¹² and concomitantly reducing the complexity of genomic DNA. Although PCR can be used to amplify small numbers of DNA molecules in low microlitre or nanolitre volumes *in vitro*, problems are associated with the dilution of heterozygous DNA. This could result in incorrect genotyping results (homozygous genotypes) because of the preferential amplification of one of the alleles. Extreme dilution indicates that there is a sufficiently high integrity of the nucleic-acid templates to allow efficient amplification. Furthermore, amplification of low concentrations of nucleic acids requires strict quality control, as it is often associated with contamination problems, which can be problematic in high-throughput applications. In contrast to nucleic-acid manipulation, it is currently impossible to amplify proteins from a biological sample.

Miniaturization in functional genomics

Genome research and functional genomics have catalysed the development of high-throughput and miniaturized approaches for the analysis of biomolecules^{5–8}.

Box 1 | Technological problems in miniaturization

Every molecular biological method can be miniaturized from microlitre to nanolitre volumes, or from molecules in the micromolar range down to single molecules (the low zeptomole range). However, miniaturization of biological assays is more complex than just transferring reactions to smaller volumes¹²². The surface to volume ratio becomes higher through miniaturization, making surface treatment important in preventing non-specific binding of components to the system walls. Formats that have a quasi-spherical shape, such as droplets, offer the best (lowest possible) surface to volume ratio, as they reduce contact with other surfaces. Further problems include the sticking of biomolecules, such as DNA or proteins, to the system wall, and even shearing of these molecules. The main structured substrate types in miniaturization comprise microarrays, nanowell plates, microchannels, nanomembranes and nanocapsules¹²³. Here, we focus on problems with microcapillary devices (microchannels) that are formed in MONOLITHIC SUBSTRATES. In these capillaries, which are normally 10–50 μm deep and 10–400 μm wide, increased flow resistance for pressured liquid flows is observed. Problems caused by the interfaces between pumps and capillaries¹²⁴ can be overcome with the use of electrokinetics, which applies electrical fields instead of using pumps and valves. Flow rates of microfluidic systems require suitable compound-mixing procedures. Also, by reducing the volume of an assay, the influence of small molecular differences increases. Reactions within small volumes will be limited by the diffusion of the components (diffusion limit). This problem can be overcome by active mixing of the liquid; for example, by surface acoustic waves¹²⁵. The important issues in applying miniaturized assays include the degree of integration of hardware components and the degree of automation that is achievable. Droplets that contain less than 100 nl of liquid can be effectively dispensed by nanoliquid handling. A decrease from 10 nl down to 50 pl can be achieved using piezopipetting systems. About 100-fl resolution is possible with bubble-jet dispensing methods. However, mainly owing to lack of integration, these potentially powerful devices are still being applied manually or with low throughput.

Microfluidic devices have mainly been used for single-step analysis¹²². Multiple-step analysis can be performed by 'lab-on-a-chip' devices that allow the integration of pumps, valves and channels on a chip. Multiple operations such as cell sorting, DNA purification and single-cell expression analysis can be carried out simultaneously^{126,127}. Interfaces between microfluidic arrays and electrospray ionization mass spectrometers have been developed for protein analysis^{128,129}. Protein identification could be achieved by using immobilized trypsin gel beads, integrated into an electro-osmotic flow system that is coupled with mass spectrometric detection^{130,131}.

LASER CAPTURE MICRODISSECTION

A method in which cells are cut out from a tissue sample using a laser beam.

CELL FRACTIONATION

Separation of cells by size, weight, density, and optical and bioelectrical properties.

MONOLITHIC SUBSTRATES

A substrate made from silicon or polymers that allows microfabrication of planar fluidic devices that have integrated microchannels or functional elements such as electrodes and detectors.

Table 1 | Dimensional scaling of miniaturized systems in genomics and proteomics

Scale	Volume relation (calculated to a cube)	Classification	Examples of samples and substrates
100 mm	1 l = (100 mm) ³	Macrostructures	The size of standard microplates with 96 to 1536 wells is 128 mm x 86 mm. The size of a standard microscopic slide is 75 mm x 25 mm.
10 mm	1 ml = (10 mm) ³	–	The well distance in standard microplates is 9 mm, 4.5 mm or 2.25 mm.
1 mm	1 µl = (1 mm) ³	–	This is the thickness of a standard microscopic slide. The spot distance of printed slides is approximately 0.2–0.9 mm.
100 µm	1 nl = (100 µm) ³	Microstructures	The size of nanowells and microchannels is in the range of 50–500 µm. The diameter of the orifice of dispensing nozzles is approximately 50–100 µm. The typical inner diameter of microcapillaries for electrophoresis is 20–100 µm.
10 µm	1 pl = (10 µm) ³	–	Cells have a diameter of approximately 10 µm. Photolithographic array structures have a dimension of 5–20 µm.
1 µm	1 fl = (1 µm) ³	–	The wavelength of visible light is in the range of 0.4–0.7 µm. The wavelength for near infrared light is 1 µm. In ESI MS, the spraying orifice has a diameter of 1–2 µm.
100 nm	n/a	Nanostructures	This is the approximate size of viruses. The typical pore diameter of nanoporous membranes in metal is 20–100 nm.
10 nm	n/a	–	This is the approximate size of macromolecules. The typical diameter of quantum dots (nanocrystals) is 2–10 nm (REF. 141), which is the detection range of atomic-force microscopy.
1 nm	n/a	–	This is the approximate size of molecules. The average diameter of carbon nanotubes is 1–2 nm (REF. 142). The diameter of nanochannels for DNA sequencing is 2 nm.
100 pm	n/a	Subnanostructures	This is the approximate size of atoms.

ESI MS, electrospray ionization mass spectrometry.

High-throughput methods to investigate DNA variation.

Mutations can significantly influence protein function and disease predisposition. Chromosomal imbalances have for a long time been studied by comparative genome hybridization (CGH) of metaphase chromosomes¹⁵. More recently, matrix CGH¹⁵, in which chromosome targets are replaced by an ordered set of specific nucleic-acid target sequences on a microarray, has been introduced as a diagnostic tool to analyse genomic alterations at a higher resolution; for example, to detect changes in gene copy number.

SNPs¹⁶ and polymorphic MICROSATELLITES¹⁷ can be used as genetic markers for genomic regions that are associated or linked to a specific trait or a common disease. Genome-wide sequencing would be the most straightforward approach to test rare and common variants for roles in complex traits¹⁸. The ultimate goal of sequencing technology is to sequence genomes of individual patients and organisms in a reasonable time (1 day) and at a low cost (~1,000 euro per human genome). Although sequencing procedures such as the Sanger method have been refined to a large extent¹⁹, current costs are still too high to make sequencing human genomes affordable. Real progress has

not been made in the past several years, despite the development of new approaches, such as using high levels of parallelization in sample preparation and using DNA MICROARRAYS or capillary devices²⁰. A new method that integrates DNA isolation, amplification and sequencing makes use of polymerase colonies (termed 'colonies')^{21,22}. Small amounts of genomic DNA are integrated into a polyacrylamide gel containing PCR reagents that covers a microscope slide. The concentration of DNA is low, so that the chromosome fragments are efficiently separated before amplification and dideoxynucleotide sequencing that is based on fluorescence detection. Multiple colonies can be sequenced in parallel *in situ*. However, this method would benefit from a higher density of colonies to accelerate throughput. Pyrosequencing is an alternative DNA-sequencing procedure for sequencing short stretches of DNA. It is useful for quantifying alleles in mixtures of DNA samples, and in small genomic regions of high sequence variation²³.

Single DNA molecule sequencing approaches have also been developed that use nanopores consisting of channels with diameters of ~2 nm (the diameter of single-stranded nucleic acids) embedded, for example, in an Si₃N₄ membrane^{24,25}.

MICROSATELLITE

A class of repetitive DNA sequences that are made up of tandemly organized repeats that are 2–8 nucleotides in length. They can be highly polymorphic and are frequently used as molecular markers in population genetics studies.

DNA MICROARRAY

An array of PCR products or oligonucleotides (corresponding to either genomic or cDNA sequences) that are deposited on solid glass slides and can be used to interrogate complex nucleic-acid samples by hybridization.

MASS SPECTROMETRY

An analytical technique for the determination of molecular mass. Although they vary greatly in design, all mass spectrometers share three general components: an ion source in which gas-phase molecular ions are produced from the analyte molecules, a mass analyser in which electrical and/or magnetic fields are used to separate the analyte ions by their different mass-to-charge (m/z) ratios and a detector for recording the separated ions.

HAPLOTYPE

The combination of alleles or genetic markers that is found on a single chromosome of a given individual.

ROLLING-CIRCLE AMPLIFICATION

A mode of DNA replication used by circular DNAs, which generates molecules that look similar to lariats. It was traditionally associated with certain bacterial plasmids and viruses, but is increasingly used as an alternative method for DNA amplification.

Unfortunately, parallelization and the production of robust membranes are important problems that need to be solved before these approaches can be applied.

Powerful methods for high-throughput SNP typing have recently been developed. They are based on the parallel generation of products that contain up to ~100,000 SNPs (this will probably increase to 1,000,000 in the next 2 years) in patient DNA, which lie in the microlitre range. These products are subsequently analysed by fluorescence analysis on DNA microarrays^{26,27} or bead-based arrays²⁸. Such procedures are particularly useful for whole-genome scans or when several thousand SNPs have to be simultaneously genotyped for large-scale approaches such as the **International HapMap Project**²⁹. 5' Nuclease (TaqMan) assays³⁰ and miniaturized MASS SPECTROMETRY procedures^{31–34} are commonly used for genotyping up to ~100 SNPs per genome in more than ~1,000 DNA samples in candidate gene studies. Furthermore, mass spectrometry can be used in the multiplex molecular analysis of HAPLOTYPES using template-DNA dilution and single-molecule amplification³⁵, allele-specific PCR³⁶ or a clone-pooling approach³⁷. The first two procedures are more useful for haplotyping short stretches of the genome in large DNA cohorts, whereas the last one is more powerful for analysing many genes or large genomic regions in a small cohort.

Recently, Landegren and co-workers have introduced a highly specific *in situ* SNP genotyping method, which makes use of allele-specific ligation of single oligonucleotide ligation probes (padlock probes) and ROLLING-CIRCLE AMPLIFICATION (RCA)³⁸. This

approach can detect single nucleotide position differences between individual nucleic acids and it has been used to characterize genetic variation in mitochondrial genomes³⁸. As this method can efficiently discriminate pools of only slightly diverse nucleic acids, it can be used for the diagnosis of bacterial RNA in biological or clinical samples.

The systematic and sensitive analysis of DNA methylation is important for understanding regulation at the DNA level^{39,40}. Analysis of genomic DNA that is treated with bisulphite, in which the non-methylated cytosines are converted to uracil while the methylated cytosines remain stable, is a common procedure for determining methylation patterns in chromosomal DNA⁴¹. High levels of sensitivity can be achieved through bisulphite treatment and subsequent PCR performed on material embedded in agarose beads⁴¹. Using this procedure, sequences of single-copy genes that are amplified from as little as 50 pg of bisulphite-modified chromosomal DNA, which is equivalent to about 10 individual cells, can be analysed with good reproducibility^{41–43}. Again, sensitivity is crucial here, as the amount of samples, especially in clinical diagnosis, is limited. A relevant application in this respect is the specific methylation profiling of gene promoters that are used for early-stage diagnosis of diseases such as cancer.

New RNA-based approaches to monitoring gene expression. Currently, one of the most frequently applied methods for studying biological networks is the comprehensive investigation of RNA expression in a cell. Expression profiles of healthy and cancerous cells can reveal meaningful changes, provide diagnostic markers and indicate potential new drug targets^{44,45}. The identification of non-coding RNAs as key players in various regulatory processes indicates that such RNAs might regulate one third of human genes^{2,46}. A detailed identification of these components and their role in regulatory networks could constitute a great challenge in itself.

RNA-expression analysis mainly builds on DNA microarrays⁴⁷, using TILING OLIGONUCLEOTIDE MICROARRAYS. However, the specificity of such microarrays with regard to allele-specific and splice-variant-specific expression profiles and the quantification capability of RNA expression is poor. Massively parallel signature sequencing (MPSS) is a new array-based technology for in-depth gene-expression profiling⁴⁸. This technology combines non-gel-based sequencing of short 'signature' sequences with the cloning of millions of DNA templates on microbeads. A planar array of at least a million of these microbeads is assembled in a flow cell at a density of roughly several million microbeads per cm². DNA sequences at the free ends of the cloned templates are simultaneously analysed by applying a fluorescence-based signature-sequencing method. Even a few mRNAs per cell can be routinely detected by this method.

Approaches to precisely monitoring gene expression at a notably reduced cost, and at a higher throughput, would significantly accelerate RNA

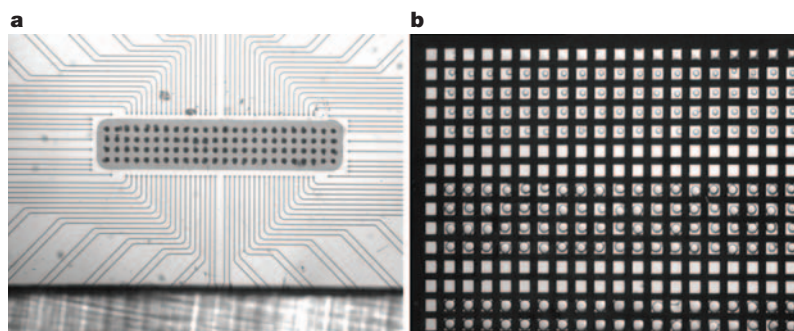


Figure 2 | Nanowell plates. a | Nanodispensing provides contact-free distribution of nanolitre droplets on nanowell plates (and microfluidic chips) without disturbing or changing the surface properties. The deposition of a nanodroplet grid onto a glass surface is carried out using the multichannel print head TopSpot (IMTEK). The volume of the droplet is 1 nl. Ninety-six droplets are arrayed in a 4 × 24 format. The print head, which has microchannels for liquid transport from reservoirs to the nozzles, is placed at a 500- μ m distance above a glass substrate in the dispensing regime. The print head is connected to a precise positioning table (400 × 400 mm²) to allow serial dispensing in requested positions and it can be adjusted in all directions and angles. The dispensing procedure is monitored with a camera from the bottom of a transparent plate. **b** | In high-density nanowell plates, the distance between neighbouring nanowells is 500 μ m in order to achieve an exact overlapping with the nozzle grid of the TopSpot print head. Nanowell plates that have well diameters of 100–500 μ m and volumes of about 1–100 nl can be designed to separate samples and reduce evaporation¹³⁹. In this figure, the wells at the top are filled with 1-nl droplets, the wells in the middle are filled with 4-nl droplets, and the wells at the bottom are filled with 8-nl droplets. Nanowell plates are made of silicon elastomers such as polymethylsiloxane, or are etched on glass slides. Nanowell chips are well-suited for miniaturized protein assays. Cell-free transcription and translation of proteins can be carried out using piezoelectrical dispensers to transfer proteins into wells, down to volumes of 100 nl, and fluorescence can be detected using a LASER SCANNING CONFOCAL MICROSCOPE¹⁴⁰.

expression profiling studies. However, sensitivity is a crucial factor here, particularly because of the restricted amounts of starting materials, such as clinical samples. This problem can be solved by using circularizable oligonucleotide probes combined with RCA of these probes⁴⁹. Alternatively, accurate and quantitative real-time reverse-transcription PCR, which is mainly used to validate microarray data⁵⁰, can be used cost effectively on nanowell plates (FIG. 2). New instruments that allow an efficient combination of liquid handling with nanowells and microfluidics are currently under development.

High-throughput functional characterization of genes can be carried out on a cell-wide level using transfected-cell microarrays⁵¹. Full-length genes cloned in expression vectors are spotted onto a glass slide with a lipid transfection reagent using an arrayer. The glass slides are covered with a layer of adherent cells. Only the cells that grow on top of the DNA spots can be transfected, so that specific genes are expressed in spatially distinct groups of cells. The phenotypic results of this 'reverse transfection' of thousands of genes can be analysed using specific cell-based assays. This approach has frequently been applied for using small interfering RNAs to silence genes⁵². Some of the limitations of this method include the cationic transfection agent used, which restricts application to a particular subset of cell types, and the difficulties in long-term storage of small interfering RNA microarrays.

Miniaturized technologies in proteomics

The systematic study of protein structures, post-translational modifications, protein profiles, protein-protein, protein-nucleic acid and protein-small molecule interactions, and the spatial and temporal expression of proteins in eukaryotic cells are crucial to understanding complex biological phenomena. Here, we describe a selection of techniques that we consider particularly useful or promising for the functional analysis of proteins.

Structural genomics. The ultimate goal of structural genomics is to determine the structure of almost all the proteins in a cell or an organism. Three-dimensional structures are important for the functional analysis of proteins in a cell and for rational drug design⁵³. In principle, databases of genome sequences allow cloning, expression and purification of all encoded proteins. DNA-cloning procedures and cDNA-clone resources, generated during the human genome sequencing era, allow at least the semi-automatic large-scale production of proteins⁵⁴. Automation of sample-purification methods significantly increased the high-throughput production of proteins. In addition, miniaturized procedures for the crystallization of proteins in 1,536-well formats or in capillaries have significantly accelerated X-ray structure analyses.

At present, it seems difficult to determine the structure of all the proteins in a cell or organism in order to understand protein function and protein-protein

interactions at the molecular level. The number of different protein variants that are due to alternative splicing and/or post-translational modifications is significantly greater than the number of genes in the human genome. Another problem is associated with the preparation of proteins for X-ray structure analysis. According to the **Ensembl** database, 22% of the proteins encoded in the human genome have at least one hydrophobic transmembrane domain, and this poses particular challenges for expression and analysis by X-RAY CRYSTALLOGRAPHY or nuclear magnetic resonance (NMR)⁵⁵. In addition, in our hands, some muscle proteins (titins) are difficult to detect due to their large size and because of solubility problems. However, these problems do not only affect structural analysis; they also influence many other aspects of protein analysis, as described in the next sections.

CRYO-ELECTRON MICROSCOPY has greatly improved our understanding of the structure and function of large native protein complexes, especially in the fields of translation, viral fusion and protein folding. It has been used for single-particle analysis or for cryo-tomography of cell organelles and protein complexes^{56,57}; techniques that have improved dramatically in terms of automated data acquisition and processing⁵⁸. One of the rate-limiting steps in single-particle analysis involves the preparation of structurally homogeneous samples. This is a great challenge as protein complexes can be structurally heterogeneous owing to the variety of post-translational modifications and binding partners. The lack of technology for such complex analysis is still a hurdle to truly high-throughput experiments in this area. X-ray structure analysis and cryo-electron microscopy are highly complementary techniques, and the mapping of high-resolution X-ray maps onto cryo-electron microscopy images will provide detailed information both on atomic and also macromolecular assemblies.

Post-translational modifications and protein profiles.

Mass spectrometry in combination with chromatographic or electrophoretic separation techniques is currently the method of choice for identifying endogenous proteins in cells, characterizing post-translational modifications and determining protein abundance⁵⁹.

Two-dimensional gel electrophoresis is unique with respect to the large number of proteins (>10,000) that can be separated and visualized in a single experiment⁶⁰. Protein spots are cut from the gel, followed by proteolytic digestion, and proteins are then identified using mass spectrometry⁶¹. Differential fluorescence labelling techniques have improved the accuracy of quantitative comparisons of protein abundance between samples⁶². However, standardization and automation of two-dimensional gel electrophoresis has proved difficult and the use of the resulting protein patterns as proteomic reference maps has only been successful in a few cases. A complementary technique, liquid chromatography, is easier to automate, and it can be directly coupled to mass spectrometry^{63,64}.

TILING OLIGONUCLEOTIDE MICROARRAYS

These microarrays contain a set of overlapping oligonucleotides that span either the entire genome, or for a more specialized approach, a subregion of interest.

LASER SCANNING CONFOCAL MICROSCOPE

A light microscope that allows imaging of fluorescent structures in thick (tens to hundreds of micrometres) specimens. A series of optical 'slices' are collected using a scanning laser beam and specially designed optics to eliminate out-of-focus excited fluorescence. The slices are reconstructed to provide detailed three-dimensional representations of the image data.

X-RAY CRYSTALLOGRAPHY

The study of the molecular structure of crystalline compounds through X-ray diffraction techniques. When an X-ray beam bombards a crystal, the atomic structure of the crystal causes the beam to scatter (diffract) in a specific pattern. X-ray crystallography provides information on the positions of individual atoms in the crystal, the distances between them, the angles of the atomic bonds and other features of molecular geometry.

CRYO-ELECTRON MICROSCOPY

Specimens are quick-frozen by plunging them into liquid ethane and are then kept frozen throughout the imaging process. Rapid freezing causes the water to form vitreous ice around the sample, preserving its native structure. Analysis of the structure can then be carried out in this state by electron microscopy.

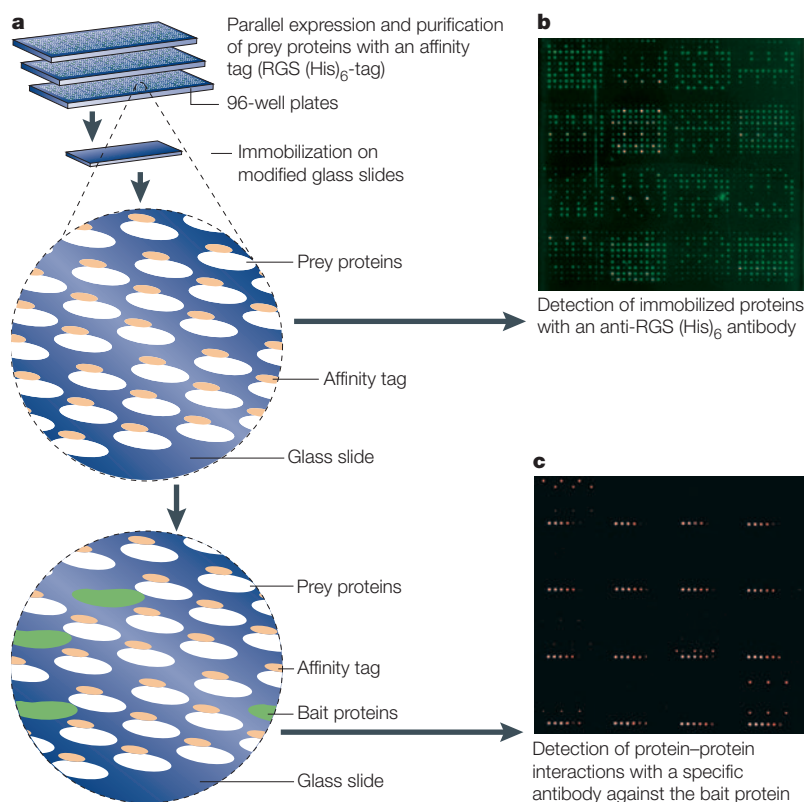


Figure 3 | Protein microarrays. a | Detecting protein–protein interactions using protein microarrays. Proteins from a ‘prey’ library (white) are expressed and affinity-purified under native conditions in a 96-well format. The purified proteins are spotted onto modified glass slides. Immobilized proteins can be visualized and quantified using an antibody that is directed against the affinity tag (beige). Alternatively, prey proteins can be incubated with a putative interaction partner (‘bait’ protein; green). After several washing steps, interactions between the bait protein and immobilized prey proteins can be detected by using a specific antibody that is directed against the bait protein. **b** | Detection of approximately 384 different (His)₆ (a histidine tag consisting of 6 amino acids)-tagged human prey proteins with an anti-RGS (His)₆ antibody. All proteins were spotted in triplicates. The signal intensity reflects the varying concentration of the spotted proteins. **c** | Visualization of putative interaction partners on a protein microarray that consists of 384 (His)₆-tagged proteins. The putative interaction partners were detected with anti-GST antibodies. White spots in **b** and **c** are protein spots that show fluorescence intensities that are greater than the dynamic range used.

MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY (MALDI MS)

An ionization technique that is suitable for mass spectrometric analysis of large biomolecules. Samples are prepared by embedding analyte molecules in a condensed matrix of small molecules. A brief laser pulse irradiates a spot on the sample, resulting in ablation of a small volume of the matrix and desorption of the embedded analyte molecules. Analyte molecules are then ionized, which leads to the formation of predominantly singly charged positive and negative analyte ions.

Minimized surface areas of nanoliquid chromatographical systems reduce analyte losses due to surface adsorption and dead volumes (the amount of overflow left in the tubing after dispensing is complete). The resolving power necessary for isolation of proteins and peptides can often be achieved by coupling two or more chromatographical separation modes in series. However, this is at the cost of increased analysis times, analyte recovery and reproducibility.

Methods that are based on stable-isotope labelling of proteins and peptides allow (relative) quantitative analysis through mass spectrometry with high accuracy⁶⁵. Numerous schemes for using stable-isotope labelling have been devised, either in combination with affinity tagging to reduce the complexity of proteomic extracts⁶⁶, or in conjunction with systemic perturbations of cultured cells to monitor the resulting alterations in protein abundance⁶⁷.

In microbiology, MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY (MALDI MS) can be used

to analyse samples from whole viruses, bacteria and spores⁶⁸. Although the mass spectra acquired from such samples are complex, a characteristic fingerprint can be created for each microorganism. In terms of clinical relevance, MALDI MS might be applied to detect biomarkers for the taxonomic identification of bacteria. Protein profiles detected by mass spectrometry represent new diagnostic approaches that can be applied, for example, to detect various stages of cancer, as cancer cells show different antigen-expression profiles to those seen in normal cells^{69,70}.

Protein–protein interactions. Affinity-based proteomic methods that are based on arrays are an alternative approach for protein profiling and can also be used to detect protein–protein interactions⁷¹. Low-density antibody microarrays, also known as analytical microarrays, have been developed to monitor the levels of marker proteins in blood and sera. Microarrays with up to 146 antibodies have been used to detect expression changes in LoVo colon carcinoma cells that are treated with ionizing radiation⁷². This procedure has revealed upregulation of apoptotic regulators induced by radiation and simultaneous downregulation of the carcino-embryonic antigen, a prototypic cancer biomarker. Recently, we analysed sera from autoimmune patients using dense protein microarrays⁷³. By comparing the pattern from patients with autoimmune disease versus healthy controls, several new antibody–autoantigen pairs were identified.

Functional microarrays are used to examine the function of uncharacterized proteins (FIG. 3). Because there are ~25,000 human genes, and even more proteins, at least 25,000 arrays spotted with 25,000 proteins would have to be screened in the search for all possible interactions. Database information can help to decrease the number of proteins to be studied for interactions on functional microarrays⁷⁴. This information, combined with bioinformatic tools, can significantly reduce the amount of *in vitro* screening by *in silico* prediction of biomolecule interactions.

A remaining problem in the application of protein microarrays is binding specificity^{75,76}, because cross-reactivity cannot accurately be predicted. Furthermore, the optimization of surface chemistry, spotting procedures, and buffer and storage conditions are still an important issue in protein microarray development⁷⁷. In addition, none of the currently used methods allow the quantitative detection of binding proteins, and it remains unclear to what extent the observed interactions are likely to represent physiological protein–protein interactions. Such information, however, is essential for algorithmic modelling of biological pathways. Detailed discussions of state-of-the-art protein arrays are provided in the articles by Seong and Choi⁷⁴, and by Espina *et al.*⁷⁸.

Procedures such as YEAST TWO-HYBRID (Y2H) ASSAYS^{79–81} and TANDEM AFFINITY PURIFICATION (TAP)^{82–84} are complementary, powerful large-scale approaches for detecting protein–protein interactions *in vivo*. However, although the Y2H technology has been automated

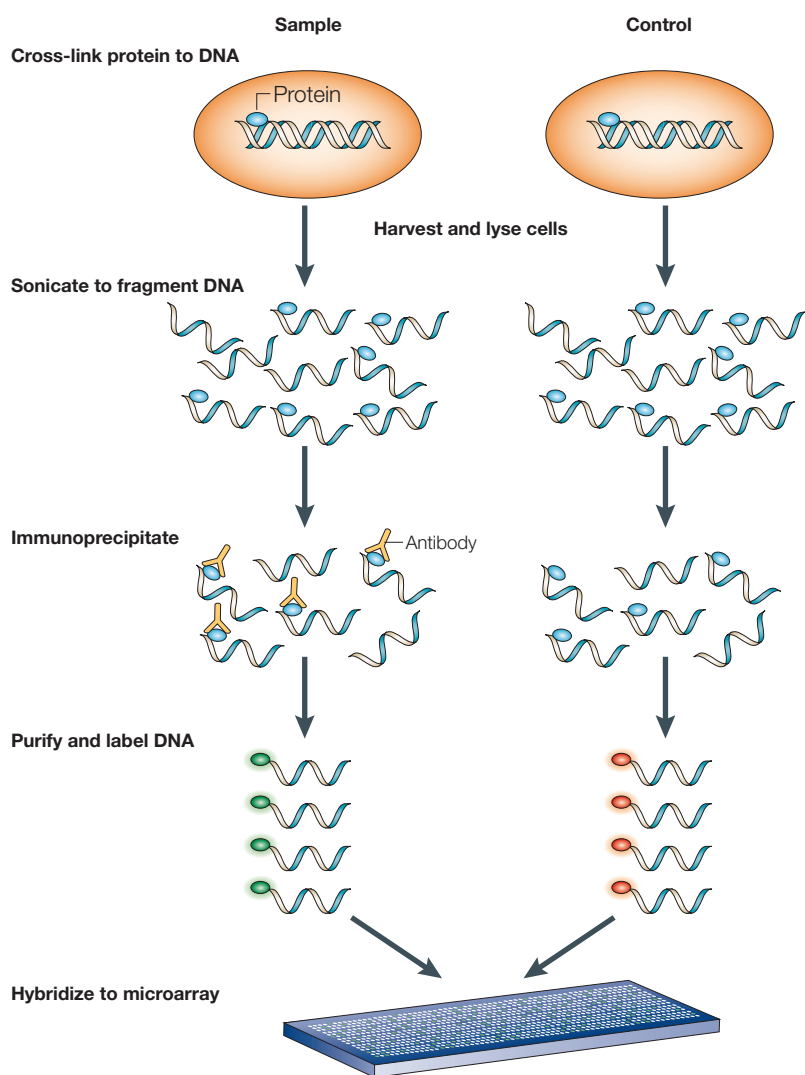


Figure 4 | Identification of protein–DNA interactions using the ChIP-chip approach. Transcription factors are cross-linked *in vivo* to their binding sites, sonicated and DNA fragments that are covalently bound to a transcription factor are enriched and purified by immunoprecipitation. The isolated DNA is subsequently tagged by fluorescence labels and hybridized on a DNA microarray, thereby allowing the identification of genomic regulatory DNA elements. Control experiments are carried out to detect non-specific background.

YEAST TWO-HYBRID (Y2H) ASSAY

One protein is fused to a transcriptional activation domain (the GAL4 activation domain) and the other to a DNA-binding domain (the GAL4 DNA-binding domain), and both fusion proteins are introduced into yeast. Expression of a (GAL4-regulated) reporter gene with the appropriate DNA-binding sites upstream of the promoter indicates that the two proteins physically interact.

and widely applied⁸⁵, the relatively new TAP technology remains to be developed, especially in the field of mammalian complex analysis. It has been shown that up to 50% of the interactions detected by Y2H can be false positives, which might be explained by interactions between proteins that do not occur *in vivo*⁷¹. As for Y2H, some comparative studies have shown that the data that are generated by the TAP method can also produce false-positive results; however, this occurs to a lesser extent (below 50%)⁸⁶.

Protein–DNA interactions. Several *in vitro* methods for detecting protein–DNA interactions have been established, including ELECTROMOBILITY SHIFT ASSAYS, ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)-based methods or phage-display approaches, which have led to the identification of several interaction

partners for one target⁸⁷. Phage display has emerged as a powerful tool for selecting a protein that recognizes the DNA target from among millions of protein variants⁸⁸. However, such *in vitro* selection methods are time consuming, have a limited throughput and often only the strongest interaction is selected. These problems have led to the development of different microarray approaches that allow high-throughput analysis.

One of the first studies to develop different microarray approaches described a universal protein–array system to study protein–DNA interactions⁸⁹. This system, using 48 different proteins, has been applied to investigating the interaction of several promoter-operator regions with various regulatory proteins^{89–91}. Recently, Snyder and co-workers used dense protein microarrays that contain nearly all the yeast proteins to identify dsDNA and ssDNA binding proteins⁹². This approach revealed several novel proteins that have DNA-binding activity and also led to the identification of a metabolic enzyme that directly regulates gene expression⁹².

A second approach to studying protein–DNA interactions was developed by Church and co-workers, who created dsDNA arrays by enzymatically converting ssDNA arrays⁹³. This approach was used to explore the DNA-binding specificities of different transcription factors at a genome- and proteome-wide level. So far, this method yields comparable results to the classical electromobility shift assays, although at a much higher throughput^{94,95}. Apart from identifying known binding sites, several putative new target regions were identified⁹⁶. Preliminary experiments using prokaryotic DNA-binding proteins in our laboratory show that identical protein–DNA complexes are identified by using protein or dsDNA microarrays (Claus Hultschig and H.S., unpublished observation). DNA microarrays are effectively employed for screening nuclear cell extracts for DNA-binding complexes, whereas protein microarrays are mainly used for identifying unknown DNA-binding proteins on a proteome-wide level. We expect that these two techniques will reveal detailed insights into transcriptional regulatory networks in the future.

The new ChIP-chip technique allows the identification of protein–DNA interactions *in vivo* (FIG. 4). It combines chromatin immunoprecipitation (ChIP) with DNA microarray (chip) detection⁹⁷. Originally developed to study protein–DNA interactions in yeast, this technique was rapidly adopted to eukaryotic cell lines and tissue samples. The procedure requires the optimization of several experimental steps, including conditions for cross-linking DNA-binding proteins to their target regions and for sonification to obtain DNA fragments of roughly homogenous lengths. ChIP-chip depends on the accessibility of at least two resources; first, a set of highly specific antibodies that are directed against the proteins of interest and second, a DNA microarray that contains the entire genome of an organism, or at least all the intergenic regions.

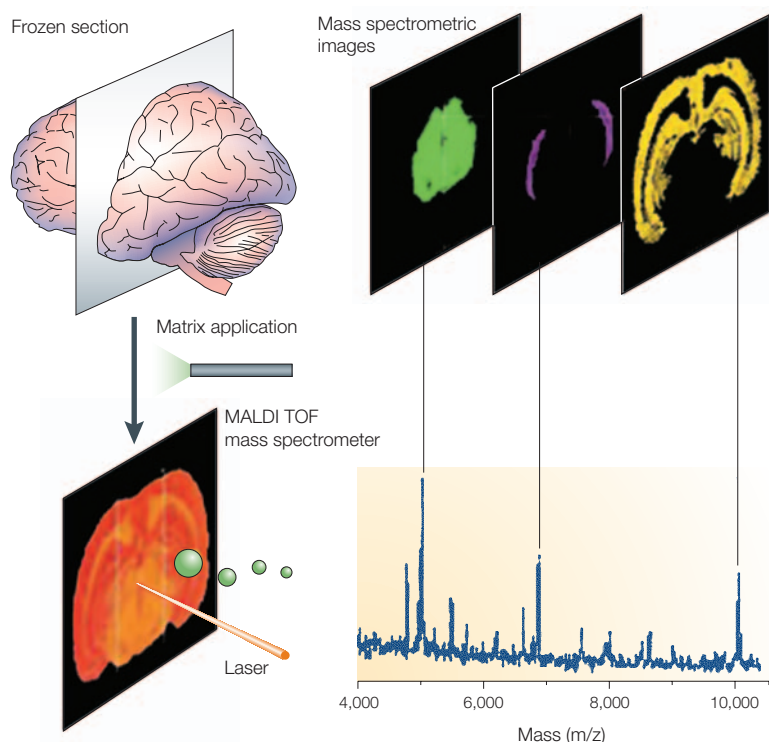


Figure 5 | The principle of MALDI imaging. Thin slices of tissue samples are attached to the MALDI (matrix-assisted laser desorption/ionization)-sample plate and coated with a MALDI matrix. A MALDI-time-of-flight (TOF) instrument is then used to acquire mass spectra in a dense pattern on the sample surface. The spatial resolution, which is limited by the diameter of the desorption/ionization laser, approaches 1 pixel per μm . Similar to the way that a picture produced by an optical scanner consists of pixels, each of which is assigned an intensity value, the mass spectrometric image also consists of pixels, each of which represents one mass spectrum. By assigning different colours to selected mass (m)-to-charge (z) values, the mass spectrometric image can be used to visualize the spatial distribution of selected compounds in the tissue studied. Modified, with permission, from *Nature Medicine* REF. 113 © (2001) Macmillan Magazines Ltd.

TANDEM AFFINITY PURIFICATION

A method that uses the expression of fusion proteins in cells carrying a double tag that is applied in two consecutive steps of purification. This protocol has the benefit of high levels of purification and native protein elution for subsequent functional, structural or biochemical analysis.

ELECTROMOBILITY-SHIFT ASSAY

An assay in which proteins that bind to a DNA fragment are detected by virtue of their reduced migration in an electrical field. The assay is often used to detect transcription-factor binding.

Protein–small molecule interactions. The use of small chemical compounds and natural products is an important step, both for pathway mapping and drug discovery⁹⁸. In principle, microarrays provide a platform for parallel and fast interaction screening, which can be performed with a minimum consumption of compounds, while still giving a high signal-to-background ratio. In early attempts, small molecules were spotted on microarrays to detect binding proteins⁹⁹. Problems that were associated with steric hindrance have now been solved by using linkers that are attached to the small molecules. Kuruvilla *et al.* have systematically screened 3,780 1,3-dioxane compounds to examine the function of the yeast protein urease β -subunit (Ure2)¹⁰⁰. The compound that was identified to specifically bind the Ure2 protein, uretupamine, is a central regulator of the nitrogen metabolic pathway.

A new approach uses soluble small molecules that are attached to a PEPTIDE NUCLEIC ACID (PNA) containing a fluorescent dye¹⁰¹. The PNA encodes a sequence that is complementary to a specific oligonucleotide, which is spotted on a DNA microarray. Proteins are mixed in solution with the PNA-modified small

molecules, followed by SIZE-EXCLUSION CHROMATOGRAPHY to separate the different complexes. The protein–small molecule complex is hybridized to a DNA microarray using a PNA tag, and quantification of the binding process is achieved through fluorescent detection. This method has been applied to monitoring protease activities in crude cell lysates and in blood samples¹⁰².

Recently, workers from Novartis Pharma AG have developed SpeedScreen, a high throughput procedure for drug discovery that is based on ELECTROSPRAY IONIZATION (ESI) mass spectrometry¹⁰³. This method provides direct identification of small molecules that bind a protein. Several hundred compounds are mixed with a purified protein, followed by size-exclusion separation in 96-well formats. The complex is then denatured and the binding compound can be accurately determined by its mass. The method was validated with known protein–ligand complexes and optimized for selection of molecules that bind with high-affinity. The liquid-handling steps and detection are robust and widely automated. A minimum of tens of thousands of compounds can be screened per day on an industrial level, using only one mass spectrometer. Further miniaturization of the liquid-handling steps to lower microlitre volumes could reduce the cost-intensive consumption of chemical compounds.

Miniaturized enzymatic assays. Protein microarrays that are based on microscope slides can be used to test enzymatic activities and for pathway mapping and substrate identification. For example, Snyder and co-workers have immobilized a set of yeast proteins with predicted kinase activity using nanowell plates¹⁰⁴. The large-scale analysis of more than 100 kinases allowed correlation of functional specificity of kinases with certain amino-acid sequences¹⁰⁴. This approach identified novel (*in vitro*) substrate specificity of kinases and has provided new data on enzyme–substrate interactions at a molecular level. Recently, we have introduced multiplex enzymatic assays on microarrays using the multiple-spotting technology, and we were able to detect 35 enzyme molecules per spot¹⁰⁵. This approach allows the rapid screening of thousands of samples on a single microarray, which has applications in drug screening and high-throughput enzymatics. Extreme sensitivity and even higher throughput might be achieved by using microfabricated substrates that enable single molecule enzymology¹⁰⁶. It is likely that the miniaturized high-throughput enzyme-screening procedures that have been mentioned in this article will evolve and become mature technologies in the next years.

Protein localization. Systematic protein-localization approaches that raise antibodies against large parts of the proteome have successfully been used for localization and expression analysis¹⁰⁷. Although this method provides high spatial resolution and sensitive detection of target proteins, it precludes the analysis of the dynamics of protein transport and localization because it requires fixation.

A more direct approach takes advantage of GFP reporter proteins and is able to monitor protein localization through real-time detection^{108,109}. Reporter proteins can be cloned in a genomic library for protein expression¹¹⁰, and the location of the GFP-tagged proteins in a cell can be detected by fluorescence microscopy. A proteome-wide approach has already been performed in yeast by Huh *et al.*¹¹¹ who studied over 6,000 proteins using GFP reporters. In mammalian, plant and *Drosophila melanogaster* cells, systematic GFP-based localization studies are also under way¹¹². Although GFP technology is useful for the systematic characterization of protein localization and dynamics, antibody-based approaches will remain valuable tools; for example, for the study of protein expression for clinical applications.

MALDI time-of-flight (TOF) imaging mass spectrometry, as illustrated in FIG. 5, is an emerging technique for the spatial and quantitative analysis of biomolecules in cell or tissue extracts¹¹³. Modern MALDI mass spectrometers are equipped with all the necessary components for this technique including imaging software (for example, see the **Bruker Daltonics** web site in the Online links box). This technique is currently used to map the distribution of selected compounds in a tissue¹¹³. The tissues under investigation contain well-defined regions, many of which have been shown to contain subsets of proteins and peptides that occur in a characteristic distribution.

New directions in diagnostic protein detection. Proteins in biological samples cannot be amplified. The transformation of protein information to DNA

molecules, however, is a new way to amplify this information for detection. RCA, which produces a long DNA molecule that contains hundreds of copies of circular DNA sequences, can be used as a powerful approach to detect proteins at high sensitivities¹¹⁴. For example, oligonucleotides that provide templates for signal amplification by RCA can be attached to capture molecules such as antibodies.

Proximity ligation is a promising method^{115,116} that uses pairs of protein binders that are tagged with DNA strands, in close proximity to a single-target molecule or molecular complex. The free ends of these DNA molecules are ligated in a proximity-dependent reaction on a complementary template DNA, so that the captured protein can be identified by a tag sequence. The sensitivity that is due to local signal amplification is high enough to detect single reacted probe pairs by microscopy. Proximity ligation has important advantages, including the high dynamic range that is achievable in detection and the highly specific analysis that is due to the simultaneous detection of a target molecule by two or more binders. Moreover, several protein epitopes can be independently identified, which is an important step in the analysis of secondary modifications. Finally, reagents that do not bind to a target molecule generate almost no detectable non-specific background. Unfortunately, the development of suitable binder pairs has not yet been automated and represents a bottleneck in the production of new proximity-ligation probes.

Low-density electrical biochips that combine silicon technology with microfluidics and biochemistry

ENZYME-LINKED

IMMUNOSORBENT ASSAY (ELISA). A widely used immunochemical method for detecting antigens or antibodies. ELISA methods are carried out in microtitre plates and use colorimetric detection.

PEPTIDE NUCLEIC ACID

A biopolymer molecule that consists of DNA bases connected by a backbone of peptide bonds, rather than phosphodiester bonds as in natural DNA.

SIZE-EXCLUSION CHROMATOGRAPHY

This method makes use of the different chromatographical behaviour that depends on the size of molecules and partly on their shape. The extent of the size exclusion of molecules is determined by the steric hindrance of analyte molecules that is due to matrix substance. Small molecules can more or less access the interior of the matrix, whereas large molecules pass by the matrix.

ELECTROSPRAY IONIZATION

An ionization technique that is suitable for mass spectrometric analysis of large biomolecules. The sample, dissolved in an aqueous-organic solvent, is pumped through a fine capillary that is made of electrically conductive material. The high voltage that is applied results in the emission of an aerosol of charged droplets of the analyte solution. Using heated gas, the spray is directed through a series of chambers that have successively decreased pressures, which results in the formation of gas-phase analyte ions.

Box 2 | Detection devices for miniaturized systems

Indirect methods that require fluorescent labelling for detection, or direct methods such as electrochemistry and mass spectrometry, have so far mainly been used in biology¹²².

Fluorescence-based detection is used to analyse single molecules at a high geometrical resolution (in the range of ~500 nm). Although fluorescence detection on surfaces allows high throughput, owing to rapid parallel read-out, the small number of detection channels is the main limitation of this method. The simultaneous analysis of up to 8 detection channels in a laser-scanning microscope is now possible owing to the recent development of adjustable acoustic optic tunable filters (for example, from Leica Microsystems), and the use of efficient algorithms (for example, the LSM 510 META from Zeiss AG) to separate dyes that have overlapping emission spectra. This allows the parallel detection of up to 8 GFPs.

The introduction of two ionization techniques, electrospray ionization mass spectrometry^{132,133} and matrix-assisted laser desorption/ionization^{134,135} in the mid-1980s, allowed the mass analysis of large biomolecules and initiated the rapid development of mass spectrometry to become a key analytical technique in the life sciences. Developments in instrument design have reduced the detection limit for proteins and peptides to attomole levels and has also improved the accuracy of mass determination in the low ppm range. New methods for fragmentation analysis have helped to determine structural features, such as amino-acid sequences and protein post-translational modifications.

Electrical detection allows single-molecule analysis to be carried out at a high resolution¹¹⁷. Industrial standards such as large-scale production and quality control are the main advantages of these approaches. Electrochemical analysis is, in principle, cheap and does not need any mechanical adjustment, as is needed in optical detection. Optical and mass spectrometric detection require more sophisticated and expensive devices and are likely to be used as complementary methods when carrying out high-throughput experiments. Electrical devices allow for extensive miniaturization and construction of small portable instruments that might be used for sensitive and specific detection of biomarkers in the future. Emerging nanomechanical and nanoelectronic detectors can provide direct, highly sensitive, real-time analysis of genes, mRNAs and proteins^{136–138}. New fabrication methods and integration with miniaturized and automated microfluidics are prerequisites for these instruments that will probably revolutionize data analysis and acquisition over the next several years. Their application will focus on the routine quantitative analysis of single cells for protein–protein, protein–DNA and protein–small molecule binding events.



represent another approach¹¹⁷, which has been used to detect complexes that consist, for example, of proteins and small molecules. Biochips that are based on silicon technology allow high spatial resolution and direct signal transfer from a sensing element to a transducer. Silicon-based electrical biochips have been developed and use redox-recycling reactions on interdigitated ultramicroelectrodes that are arranged in array formats¹¹⁸. Binding events can be monitored by redox-recycling of products, which is generated through enzyme labelling of affinity complexes¹¹⁹. The principles of this method can be extended to other formats such as integrated 'lab-on-a-chip' devices or metal nanoparticles. Low-density electrical biochips will probably become routine instruments for diagnostics, provided that the market situation will allow their launch.

Outlook

Many biomolecules have to be analysed to investigate complex biological systems. Although a limited number of key factors of biological relevance might be sufficient to provide predictive models¹²⁰, marker molecules have to be analysed in many samples for potential diagnosis and drug development¹²¹. Real progress in the diagnosis and therapy of almost all complex diseases requires the identification and in-depth analysis of many, if not all, components that influence these biological processes.

Similar to information technology, miniaturization has revolutionized biology. New hardware components for liquid handling and sensitive detection devices have helped the development of efficient and sensitive

high-throughput methods that provide more experimental data for biological problem-solving (BOX 2). Several simple *in vitro* methods, such as oligonucleotide hybridization, have been miniaturized using microarray formats and are now routinely applied. However, new concepts are required not only for high-throughput sequencing, but also for many other purposes, which have been outlined in this article. Methods to specifically detect individual biomolecules at the level of populations and within and among cells are difficult to achieve. At present, only a few molecules can be identified and analysed at the single-molecule level in complex mixtures.

A further problem concerns data handling. Molecular data sets such as gene-expression profiles can be produced in a relatively short time. However, the lack of standardization of data between laboratories is obvious. Data standardization is required for consistent analysis of different biological data sets that are stored in databases. Currently, the first efforts are being made to structure the heterogeneous data obtained from high-throughput experiments from many laboratories. Agreements on standardization, as well as creating interconnected databases, are essential prerequisites for the efficient analysis of molecular networks. Initiatives have been launched to standardize the bioinformatic treatment of microarray data interpretation (for example, the MIAME (minimum information about a microarray experiment) and MAGE (microarray and gene expression) initiatives; see the [Microarray Gene Expression Data Society](#) web site in the Online links box), which will probably be adopted for experimentation and database handling in the future.

- Hood, L., Heath, J. R., Phelps, M. E. & Lin, B. Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643 (2004).
- Huttenhofer, A., Brosius, J. & Bachellerie, J.-P. RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.* **6**, 835–843 (2002).
- Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
A systematic investigation of complex biological networks.
- Butcher, E. C., Berg, E. L. & Kunkel, E. J. Systems biology in drug discovery. *Nature Biotechnol.* **22**, 1253–1259 (2004).
- Southern, E. M. DNA microarrays. History and overview. *Methods Mol. Biol.* **170**, 1–15 (2001).
- Lennon, G. G. & Lehrach, H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* **7**, 314–317 (1991).
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
One of the first papers that shows comprehensive gene-expression analysis of microarrays.
- Fodor, S. P. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
The construction of high-density oligonucleotide microarrays is shown.
- Bussow, K. *et al.* A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* **26**, 5007–5008 (1998).
One of the first protein-array papers.
- Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
Large-scale functional analysis of proteins on microarrays.
- Church, J. G., Stapleton, E. A. & Reilly, B. D. Isolation of high quality mRNA from a discrete cell cycle population identified using a nonvital dye and fluorescence activated sorting. *Cytometry* **14**, 271–275 (1993).
- Whetsell, L., Maw, G., Nadon, N., Ringer, D. P. & Schaefer, F. V. Polymerase chain reaction microanalysis of tumors from stained histological slides. *Oncogene* **7**, 2355–2361 (1992).
- Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R. & Liotta, L. A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet.* **14**, 272–276 (1998).
- Dreger, M., Bengtsson, L., Schoneberg, T., Otto, H. & Hucho, F. Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc. Natl Acad. Sci. USA* **98**, 11943–11948 (2001).
- Solinias-Toldo, S. *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).
- Syvanen, A. C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Rev. Genet.* **2**, 930–942 (2001).
- Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
- Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
- Sanger, F., Nickens, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
- Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nature Rev. Genet.* **5**, 335–344 (2004).
- Mitra, R. D. & Church, G. M. *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* **27**, e34 (1999).
- Mitra, R. D., Shendure, J., Olejnik, J., Edyta Krzymanska, O. & Church, G. M. Fluorescent *in situ* sequencing on polymerase colonies. *Anal. Biochem.* **320**, 55–65 (2003).
- Pomaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
- Deamer, D. W. & Akeson, M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol.* **18**, 147–151 (2000).
- Li, J. *et al.* Ion-beam sculpting at nanometre length scales. *Nature* **412**, 166–169 (2001).
- Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnol.* **5**, 1–5 (2003).
- Matsuzaki *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* **2**, 109–111 (2005).
- Oliphant, A., Barker, D. L., Stuelpragel, J. R. & Chee, M. S. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques Suppl.* **56**–58, 60, 61 (2002).
- The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
- Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 7276–7280 (1991).
- Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl Acad. Sci. USA* **96**, 10016–10020 (1999).
- Sauer, S. *et al.* Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Res.* **30**, e22 (2002).
- Sauer, S., Lehrach, H. & Reinhardt, R. MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging. *Nucleic Acids Res.* **31**, e63 (2003).
- Sauer, S. & Gut, I. G. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **782**, 73–87 (2002).
- Ding, C. & Cantor, C. R. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. *Proc. Natl Acad. Sci. USA* **100**, 7449–7453 (2003).



36. Tost, J. *et al.* Molecular haplotyping at high throughput. *Nucleic Acids Res.* **30**, e96 (2002).
37. Burgtorf, C. *et al.* Clone-based systematic haplotyping (CSH): a procedure for physical haplotyping of whole genomes. *Genome Res.* **13**, 2717–2724 (2003).
38. Larsson, C. *et al.* *In situ* genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nature Methods* **1**, 227–232 (2004).
39. Walter, J. & Paulsen, M. Imprinting and disease. *Semin. Cell Dev. Biol.* **14**, 101–110 (2003).
40. Rakyan, V. K. *et al.* DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* **2**, e405 (2004).
41. Olek, A., Oswald, J. & Walter, J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**, 5064–5066 (1996).
42. Adorjan, P. *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* **30**, e21 (2002).
43. Cottrell, S. E. *et al.* A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res.* **32**, e10 (2004).
44. Welsh, J. B. *et al.* Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**, 5974–5978 (2001).
45. Dhanasekaran, S. M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
46. Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242–2246 (2004).
47. Lander, E. S. Array of hope. *Nature Genet.* **21**, 3–4 (1999).
48. Brenner, S. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnol.* **18**, 630–634 (2000).
49. Lizardi, P. M. *et al.* Mutation detection and single molecule counting using isothermal rolling circle amplification. *Nature Genet.* **19**, 225–232 (1998).
50. Chuai, R. F. *et al.* Post-analysis follow-up and validation of microarray experiments. *Nature Genet.* **32**, 509–514 (2002).
51. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110 (2001).
52. Dorsett, Y. & Tuschl, T. siRNAs: applications in functional genomics and potential as therapeutics. *Nature Rev. Drug Discov.* **3**, 318–329 (2004).
53. Zhang, C. & Kim, S. H. Overview of structural genomics: from structure to function. *Curr. Opin. Chem. Biol.* **7**, 28–32 (2003).
54. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39–43 (2003).
55. Birney, E. *et al.* An overview of Ensembl. *Genome Res.* **14**, 925–928 (2004).
56. Baumeister, W. & Steven, A. C. Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem. Sci.* **25**, 624–631 (2000).
57. Frank, J. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 303–319 (2002).
58. Subramaniam, S. & Milne, J. L. Three-dimensional electron microscopy at molecular resolution. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 141–155 (2004).
59. Zhu, H., Bilgin, M. & Snyder, M. Proteomics. *Annu. Rev. Biochem.* **72**, 783–812 (2003).
60. Klose, J. & Kobalz, U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**, 1034–1059 (1995).
61. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
62. Seike, M. *et al.* Proteomic signature of human cancer cells. *Proteomics* **4**, 2776–2788 (2004).
63. Gaskell, S. J. Electrospray: principles and practice. *J. Mass Spectrom.* **32**, 677–688 (1997).
64. Mirgorodskaya, E., Brauer, C., Fucini, P., Lehrach, H. & Gobom, J. Nanoflow liquid chromatography coupled to matrix-assisted laser desorption/ionization mass spectrometry: sample preparation, data analysis, and application to the analysis of complex peptide mixtures. *Proteomics* **13**, 399–408 (2005).
65. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
66. Ong, S. E., Foster, L. J. & Mann, M. Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124–130 (2003).
67. Gustavsson, N. *et al.* A proteomic method for the analysis of changes in protein concentration in response to systemic perturbations using metabolic incorporation of stable isotopes and mass spectrometry. *Proteomics* (in the press).
68. Claydon, M. A., Davey, S. N., Edwards-Jones, V. & Gordon, D. B. The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotechnol.* **14**, 1584–1586 (1996).
69. Tang, N., Tomatore, P. & Weinberger, S. R. Current developments in SELDI affinity technology. *Mass Spectrom. Rev.* **23**, 34–44 (2004).
70. Qu, Y. *et al.* Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin. Chem.* **48**, 1835–1843 (2002).
71. Agaton, C. *et al.* Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell Proteomics* **2**, 405–414 (2003).
72. Sreekumar, A. *et al.* Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.* **61**, 7585–7593 (2001).
73. Lueking, A. *et al.* A nonredundant human protein chip for antibody screening and serum profiling. *Mol. Cell Proteomics* **2**, 1342–1349 (2003).
74. Seong, S. Y. & Choi, C. Y. Current status of protein chip development in terms of fabrication and application. *Proteomics* **3**, 2176–2189 (2003).
75. Michaud, G. A. *et al.* Analyzing antibody specificity with whole proteome microarrays. *Nature Biotechnol.* **21**, 1509–1512 (2003).
- A study describing the specificity problems that are associated with the analysis of large sets of proteins using antibodies.**
76. Schweitzer, B., Predki, P. & Snyder, M. Microarrays to characterize protein interactions on a whole-proteome scale. *Proteomics* **3**, 2190–2199 (2003).
77. Angenendt, P., Glöckler, J., Murphy, D., Lehrach, H. & Cahill, D. J. Toward optimized antibody microarrays: a comparison of current microarray support materials. *Anal. Biochem.* **309**, 253–260 (2002).
78. Espina, V. *et al.* Protein microarray detection strategies: focus on direct detection technologies. *J. Immunol. Methods* **290**, 121–133 (2004).
79. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
80. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
81. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574 (2001).
82. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
83. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
84. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
85. Goehler, H. *et al.* A protein interaction network links GIT1, an enhancer of Huntingtin aggregation, to Huntington's disease. *Mol. Cell* **15**, 853–865 (2004).
86. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
87. Walter, G., Konthar, Z. & Lehrach, H. High-throughput screening of surface displayed gene products. *Comb. Chem. High Throughput Screen* **4**, 193–205 (2001).
88. Cicchini, C. *et al.* Searching for DNA–protein interactions by λ -phage display. *J. Mol. Biol.* **322**, 697–706 (2002).
89. Ge, H. UPA, a universal protein array system for quantitative detection of protein–protein, protein–DNA, protein–RNA and protein–ligand interactions. *Nucleic Acids Res.* **28**, e3 (2000).
90. Snappan, M. *et al.* Dissecting DNA–protein and protein–protein interactions involved in bacterial transcriptional regulation by a sensitive protein array method combining a near-infrared fluorescence detection. *Proteomics* **3**, 647–657 (2003).
91. Kersten, B. *et al.* Protein microarray technology and ultraviolet crosslinking combined with mass spectrometry for the analysis of protein–DNA interactions. *Anal. Biochem.* **331**, 303–313 (2004).
92. Hall, D. A. *et al.* Regulation of gene expression by a metabolic enzyme. *Science* **306**, 482–484 (2004).
93. Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M. Quantifying DNA–protein interactions by double-stranded DNA arrays. *Nature Biotechnol.* **17**, 573–577 (1999).
- One of the first studies to show a DNA–protein interaction on an array.**
94. Bulyk, M. L., Huang, X., Choo, Y. & Church, G. M. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc. Natl. Acad. Sci. USA* **98**, 7158–7163 (2001).
95. Linnell, J. *et al.* Quantitative high-throughput analysis of transcription factor binding specificities. *Nucleic Acids Res.* **32**, e44 (2004).
96. Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nature Genet.* **36**, 1331–1339 (2004).
97. Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. & Farnham, P. J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.* **16**, 235–244 (2002).
98. Darvas, F. *et al.* Recent advances in chemical genomics. *Curr. Med. Chem.* **11**, 3119–3145 (2004).
99. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
- One of the first proof-of-principle studies to show the applications for protein microarrays.**
100. Kuruvilla, F. G., Shamji, A. F., Sternon, S. M., Hergenrother, P. J. & Schreiber, S. L. Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* **416**, 653–657 (2002).
101. Winssinger, N., Harris, J. L., Backes, B. J. & Schultz, P. G. From split-pool libraries to spatially addressable microarrays and its application to functional proteomic profiling. *Angew. Chem. Int. Ed. Engl.* **40**, 3152–3155 (2001).
102. Winssinger, N., Ficarro, S., Schultz, P. G. & Harris, J. L. Profiling protein function with small molecule microarrays. *Proc. Natl. Acad. Sci. USA* **99**, 11139–11144 (2002).
103. Muckenschnabel, I., Falchetto, R., Mayr, L. M. & Filipuzzi, I. SpeedScreen: label-free liquid chromatography–mass spectrometry-based high-throughput screening for the discovery of orphan protein ligands. *Anal. Biochem.* **324**, 241–249 (2004).
104. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283–289 (2000).
- This is a large-scale analysis of protein function using nanowells.**
105. Angenendt, P., Lehrach, H., Kreutzberger, J. & Glöckler, J. Subnanoliter enzymatic assays on microarrays. *Proteomics* **5**, 420–425 (2005).
106. Rondelez, Y. *et al.* Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nature Biotechnol.* **23**, 361–365 (2005).
107. Agaton, C. *et al.* Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell Proteomics* **2**, 405–414 (2003).
108. Hadjantonakis, A. K., Dickinson, M. E., Fraser, S. E. & Papaioannou, V. E. Technicolour transgenics: imaging tools for functional genomics in the mouse. *Nature Rev. Genet.* **4**, 613–625 (2003).
109. Verkhusha, V. V. & Lukyanov, K. A. The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nature Biotechnol.* **22**, 289–296 (2004).
110. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R. & Wiemann, S. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1**, 287–292 (2000).
111. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- The authors provide a systematic analysis of protein localization in yeast using GFPs.**
112. Davis, T. N. Protein localization in proteomics. *Curr. Opin. Chem. Biol.* **8**, 49–53 (2004).
113. Stoekli, M., Chaurand, P., Hallahan, D. E. & Caprioli, R. M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature Med.* **7**, 493–496 (2001).
- This is the proof-of-concept paper on MALDI imaging.**
114. Schweitzer, B. *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnol.* **20**, 359–365 (2002).
115. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnol.* **20**, 473–477 (2002).
- The authors describe a highly sensitive and specific method for the detection of proteins that elegantly combines nucleic-acid and protein chemistry.**
116. Gullberg, M. *et al.* Cytokine detection by antibody-based proximity ligation. *Proc. Natl. Acad. Sci. USA* **101**, 8420–8424 (2004).
117. Albers, J., Grunwald, T., Nebling, E., Piechotta, G. & Hintsche, R. Electrical biochip technology — a tool for microarrays and continuous monitoring. *Anal. Bioanal. Chem.* **377**, 521–527 (2003).
- An overview on current developments in electrical biochip technology.**
118. Paeschke, M., Dietrich, F., Uhlig, A. & Hintsche R. Voltammetric multichannel measurements using silicon fabricated microelectrode arrays. *Electroanalysis* **10**, 891–898 (1996).
119. Hintsche, R., Albers, J., Bernt, H. & Eder, A. Multiplexing of microelectrode arrays in voltammetric measurements. *Electroanalysis* **12**, 660–665 (2000).



120. Hood, L. & Perlmutter, R. M. The impact of systems approaches on biological problems in drug discovery. *Nature Biotechnol.* **22**, 1215–1217 (2004).
121. Weston, A. D. & Hood, L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* **3**, 179–196 (2004).
122. Bilitewski, U., Genrich, M., Kadow, S. & Mersal, G. Biochemical analysis with microfluidic systems. *Anal. Bioanal. Chem.* **377**, 556–569 (2003).
123. Whitesides, G. M. The 'right' size in nanobiotechnology. *Nature Biotechnol.* **21**, 1161–1165 (2003).
- An informative introduction to nanobiotechnology.**
124. Hong, J. W. & Quake, S. R. Integrated nanoliter systems. *Nature Biotechnol.* **21**, 1179–1183 (2003).
125. Toegl, A., Kirchner, R., Gauer, C. & Wixforth, A. Enhancing results of microarray hybridizations through microagitation. *J. Biomol. Tech.* **14**, 197–204 (2003).
126. Fu, A. Y., Chou, H. P., Spence, C., Arnold, F. H. & Quake, S. R. An integrated microfabricated cell sorter. *Anal. Chem.* **74**, 2451–2457 (2002).
127. Hong, J. W. & Quake, S. R. Integrated nanoliter systems. *Nature Biotechnol.* **21**, 1179–1183 (2003).
128. Xue, Q. *et al.* Multichannel microchip electrospray mass spectrometry. *Anal. Chem.* **69**, 426–430 (1997).
129. Figeys, D., Gygi, S. P., McKinnon, G. & Aebersold, R. An integrated microfluidics-tandem mass spectrometry system for automated protein analysis. *Anal. Chem.* **70**, 3728–3734 (1998).
130. Wang, C. *et al.* Integration of immobilized trypsin bead beds for protein digestion within a microfluidic chip incorporating capillary electrophoresis separations and an electrospray mass spectrometry interface. *Rapid Commun. Mass Spectrom.* **14**, 1377–1383 (2000).
131. Jin, L. J., Ferrance, J., Sanders, J. C. & Landers, J. P. A microchip-based proteolytic digestion system driven by electroosmotic pumping. *Lab. Chip* **3**, 11–18 (2003).
132. Yamashita, M. & Fenn, J. B. Electrospray ion source: another variation on the free-jet theme. *J. Phys. Chem.*, **88**, 4451–4459 (1984).
133. Alexandrov, M. L. Ion extraction from solutions of atmospheric pressure — a method of mass spectrometric analysis of bioorganic substances. *Dokl. Akad. Nauk. SSSR*, **277**, 379–383 (1984).
134. Karas, M. & Hillenkamp, F. Laser desorption/ionization of proteins with molecular masses exceeding 10000 daltons. *Anal. Chem.* **60**, 2299–2303 (1988).
135. Tanaka, K., Waki, H., Ido, Y., Akita, S. & Yoshida, Y. Protein and polymer analysis up to M/Zx 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).
136. Fritz, J. *et al.* Translating biomolecular recognition into nanomechanics. *Science* **288**, 316–318 (2000).
137. Cui, Y., Wei, Q., Park, H. & Lieber, C. M. Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* **293**, 1289–1292 (2001).
138. Chen, R. J. *et al.* Noncovalent functionalization of carbon nanotubes for highly specific electronic biosensors. *Proc. Natl Acad. Sci. USA* **100**, 4984–4989 (2003).
139. Biran, I. & Walt, D. R. Optical imaging fiber-based single live cell arrays: a high-density cell assay platform. *Anal. Chem.* **74**, 3046–3054 (2002).
140. Angenendt, P. *et al.* Cell-free protein expression and functional assay in nanowell chip format. *Anal. Chem.* **76**, 1844–1849 (2004).
141. Han, M., Gao, X., Su, J. Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nature Biotechnol.* **19**, 631–635 (2001).
142. Ajayan, P. M., Charlier, J.-C. & Rinzler, A. G. Carbon nanotubes: From macromolecules to nanotechnology. *Proc. Natl Acad. Sci. USA* **96**, 14199–14200 (1999).

Acknowledgements

We would like to thank W. Nietfeld for providing FIG. 1 and would like to acknowledge the European Union, the Senatsverwaltung für Wissenschaft, Forschung und Kultur, Berlin (Ultra-Structure Network), the German Ministry for Research and Education (NGFN), and the Max Planck Society for financial support.

Competing interests statement

The authors declare no competing financial interests.

 **Online links**

FURTHER INFORMATION

- Biochemical Tools Group at the Max Planck Institute:** http://www.molgen.mpg.de/~ag_sauer
- Bruker Daltonics web site:** <http://www.bdal.com>
- Ensembl:** <http://www.ensembl.org>
- GenomeMatrix Database:** <http://genome-matrix.org/>
- International HapMap Project:** <http://www.hapmap.org>
- Lehrach's homepage:** <http://www.molgen.mpg.de/research/lehrach>
- Max Planck Institute for Molecular Genetics:** <http://www.molgen.mpg.de>
- Microarray Gene Expression Data Society web site:** <http://www.mged.org>
- MolTools (Advanced Molecular Tools for Array-based Analysis of Genomes):** <http://www.moltools.org>
- Protein Structure Factory:** <http://www.proteinstrukturfabrik.de>
- RZPD — German Resource Centre for Genome Research:** <http://www.rzpd.de>
- Access to this interactive links box is free online.**