

- 7 Meltzer, P.S. *et al.* (1992) Rapid generation of region specific probes by chromosome microdissection and their application. *Nat. Genet.* 1, 24–28
- 8 Lengauer, C. *et al.* (1993) Chromosomal bar codes produced by multicolor fluorescence *in situ* hybridization with multiple YAC clones and whole chromosome painting probes. *Hum. Mol. Genet.* 2, 505–512
- 9 Müller, S. *et al.* (1997) Toward a multicolor chromosome bar code for the entire human karyotype by fluorescence *in situ* hybridization. *Hum. Genet.* 100, 271–278
- 10 Müller, S. *et al.* (1998) Cross-species colour segmenting: a novel tool in human karyotype analysis. *Cytometry* 33, 445–452
- 11 Chudoba, I. *et al.* (1999) High resolution multicolor-banding: a new technique for refined FISH analysis of human chromosomes. *Cytogenet. Cell Genet.* 84, 156–160
- 12 Mrasek, K. *et al.* (2001) Reconstruction of the female Gorilla gorilla karyotype using 25-color FISH and multicolor banding (MCB). *Cytogenet. Cell Genet.* 93, 242–248
- 13 Schröck, E. *et al.* (1996) Multicolor spectral karyotyping of human chromosomes. *Science* 273, 494–497
- 14 Speicher, M.R. *et al.* (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* 12, 368–375
- 15 Kallioniemi, A. *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–820
- 16 Schröck, E. *et al.* (1997) Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities. *Hum. Genet.* 101, 255–262
- 17 Uhrig, S. *et al.* (1999) Multiplex-FISH for pre- and postnatal diagnostic applications. *Am. J. Hum. Genet.* 65, 448–462
- 18 Tönnies, H. *et al.* (2001) Comparative genomic based strategy for the analysis of different chromosome imbalances detected in conventional cytogenetic diagnostics. *Cytogenet. Cell Genet.* 93, 188–194
- 19 Schwarzbach and Heslop-Harrison (2000) *Practical in situ Hybridization*, BIOS Scientific Publishers Limited
- 20 Solinas-Toldo, S. *et al.* (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399–407
- 21 Pinkel, D. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211
- 22 Pollack, J.R. *et al.* (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46
- 23 Snijders, A.M. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* 29, 263–264

#### Holger Tönnies

Charité, Campus-Virchow, Institute of Human Genetics, Humboldt-University Berlin, Augustenburger Platz 1, D-13353 Berlin, Germany.  
e-mail: Holger.Toennies@charite.de

## High-throughput protein arrays: prospects for molecular diagnostics

Gerald Walter, Konrad Büssow, Angelika Lueking and Jörn Glökler

**High-throughput protein arrays allow the miniaturized and parallel analysis of large numbers of diagnostic markers in complex samples. Using automated colony picking and gridding, cDNA or antibody libraries can be expressed and screened as clone arrays. Protein microarrays are constructed from recombinantly expressed, purified, and yet functional proteins, entailing a range of optimized expression systems. Antibody microarrays are becoming a robust format for expression profiling of whole genomes. Alternative systems, such as aptamer, PROfusion™, nano- and microfluidic arrays are all at proof-of-concept stage. Differential protein profiles have been used as molecular diagnostics for cancer and autoimmune diseases and might ultimately be applied to screening of high-risk and general populations.**

Published online: 8 May 2002

Protein arrays appear to be promising tools for drug screening and diagnostics [1]. The concept of the arrayed library allows gene expression analysis and protein interaction screening on a whole-genome scale [2]. In contrast to DNA arrays, gene action can be studied directly, providing the proteins' natural shape and functionality are maintained. This requires novel systems

for high-throughput protein expression, which produce sufficient amounts of properly modified and folded molecules. Large numbers of proteins must be arrayed at high density, keeping them intact and biologically active. This is most easily achieved if molecules of the same general structure (e.g. antibodies) are arrayed. Antibody arrays are now becoming an important screening tool for a wide range of molecules in complex mixtures (e.g. body fluids). Alternatively, nucleotide aptamers might be able to mimic certain protein functions and make even more robust array formats in the future. Here we discuss the main technological challenges, recent progress and diagnostic prospects of high-throughput protein arrays (Table 1), a field that is still largely experimental but rapidly progressing.

#### Protein features and expression systems

Proteins constitute a large variety of shapes and features that govern their functions. Three-dimensional protein structures are determined by amino-acid sequences and environmental conditions. Post-translational modifications influence binding and function of many eukaryotic proteins. Glycosylation affects solubility, lifetime and activity of some proteins, such as antibodies' ability to bind

complement [3]. Phosphorylation regulates enzyme activities and allows specific protein–protein interactions via domains recognizing phosphate groups (e.g. SH-2 domains that interact with phosphotyrosine groups in signal transduction [4]). Such post-translational modifications depend strongly on the expressed protein's cellular host. If recombinant proteins are expressed in bacteria, they are usually not glycosylated or phosphorylated, whereas *S. cerevisiae* tends to hyperglycosylate secreted mammalian proteins. Targeting of recombinant proteins to secretory compartments, such as the endoplasmic reticulum, allows them to form stabilizing disulfide bridges. Globular proteins consist of a hydrophilic exterior and a hydrophobic interior. If exposed to hydrophobic surfaces, organic solvents or detergents, the inside is turned out, thus rendering these proteins inactive. This emphasizes the imperative of proper handling of native proteins for functional assays and three-dimensional interaction studies. Refolding denatured proteins *in vitro* is an arduous task, requiring customized protocols and is not attainable in many cases.

Protein expression systems should be chosen carefully. *E. coli* is still the most convenient host organism, but many eukaryotic proteins end up in cytoplasmic

**Table 1. High-throughput protein array techniques and diagnostic applications**

| Techniques                          | Advantages                                                                                                                                                                           | Limitations                                                                                    | Diagnostic applications                                                                | Refs                        |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------------------|
| cDNA clone arrays                   | Automated colony picking and gridding, direct relation to clone DNA                                                                                                                  | Cellular protein background, undefined proteins or fragments                                   | Protein expression analysis, protein interaction screening                             | [10–12]                     |
| Protein microarrays                 | Most sensitive, miniaturized and parallel capture-target assays, analysis of complex samples                                                                                         | Different protein features, no standardized handling of native proteins, solid-phase chemistry | Protein expression analysis, molecular identification, quantification, affinity assays | [1,2,5,9,15,17,18,20,21,30] |
| Antibody clone arrays               | As cDNA; isolation of antibodies against impure proteins and complex antigens possible                                                                                               | Antigen-coated filters required, sensitivity of detection to be improved by 100–1000-fold      | Antibody expression, screening and selection                                           | [24]                        |
| Antibody microarrays                | As protein; robust, standard molecular structure                                                                                                                                     | Limited proteome and marker coverage by antibodies available                                   | As protein; protein marker and expression profiling                                    | [17]                        |
| Aptamer microarrays                 | Robust, standard DNA microarray technology, covalent attachment of aptamers to cognate proteins, direct protein staining, aptamer specificity optimized by <i>in vitro</i> evolution | Proof-of-concept stage                                                                         | As antibody                                                                            | [27]                        |
| PROfusion™ arrays                   | <i>In vitro</i> covalent linkage of a protein to its encoding RNA, self-assembly to surface-bound DNA probes, directed <i>in vitro</i> evolution                                     | Proof-of-concept stage                                                                         | As protein and antibody                                                                | [19]                        |
| Protein nanoarrays                  | No detectable nonspecific binding of proteins, can be screened by atomic-force microscopy                                                                                            | Proof-of-concept stage                                                                         | As protein and antibody                                                                | [28]                        |
| Microfluidic arrays (Lab-on-a-Chip) | Optimal surface-to-volume ratio, shorter hybridization times, large binding-signal capacity, readily automated, fast and robust                                                      | Proof-of-concept stage                                                                         | As protein and antibody                                                                | [16]                        |

inclusion bodies and can only be recovered in denatured form. The retained display of linear epitopes might be sufficient for antigen–antibody screening purposes, such as in autoimmune diagnostics [5]. However, if native proteins are to be expressed in *E. coli*, they have to be targeted to compartments where they fold naturally, for example by secretion into the periplasm or membrane integration via appropriate leader sequences [6]. Alternatively, and especially if post-translational modifications are required, folded proteins can be efficiently produced in yeast (e.g. *P. pastoris* [7]). Therefore, protein expression libraries should be cloned in various vectors (or shuttle vectors, see below), and protein-expressing or protein-binding clones should be re-arrayed after screening (Fig. 1) to make sure that all expressible clones find optimal host conditions and maximum protein diversity is obtained from a library.

Manufacturing high-throughput protein arrays requires the efficient production of large numbers of proteins. This is a major technical challenge, calling for highly parallel, preferably automated, recombinant expression systems.

Prokaryotic genome fragments can be directly cloned into expression vectors. In eukaryotes, the presence of non-coding regions (introns) within genes requires the cloning of cDNA from mRNA to join the coding regions (exons). The recombinant expression of all open reading frames of *S. cerevisiae* has been reported [8]. An almost-complete set of yeast strains (expressing 6144 open reading frames as fusion proteins) was generated and grown in defined pools, which were assayed for biochemical (e.g. enzyme) activities and deconvoluted to identify the active source strains. 5800 yeast proteins were individually expressed, purified and arrayed to form the first whole-genome protein microarray [9]. For high-throughput expression and analysis of unknown proteins, arrayed cDNA libraries, cloned in bacterial or yeast expression vectors, have been developed [10]. These libraries eliminate the need to construct individual expression systems for every protein of interest. By arraying clones on high-density filters, the protein products of complete libraries are screened in parallel, and expression clones of interest can be identified (Fig. 1).

The same libraries are used for high-throughput expression in microtitre plates and protein characterization by mass spectrometry [11]. Mammalian clone arrays were recently described, featuring cDNAs printed on glass slides, which were taken up by cells growing on top, creating spots of localized transfection within a lawn of nontransfected cells [12].

For parallel evaluation of different expression systems, high-throughput subcloning of human open reading frames has been described [13]. Recombination-based cloning allows a rapid exchange of vectors and expression systems, such as the Gateway system (Invitrogen, Rockville, WI, USA) or the Creator system (Clontech, Palo Alto, CA, USA). Alternatively, shuttle vectors enable parallel protein expression in different hosts (e.g. *E. coli* and *P. pastoris*) without subcloning [7]. The pTriEx vector (Novagen, Madison, WI, USA) can be used for expression in *E. coli*, baculovirus-infected insect cells and vertebrate cells. *In vitro* transcription–translation systems seem now to provide small, yet sufficient, amounts of protein for microarray production. A combination of *in vitro* protein

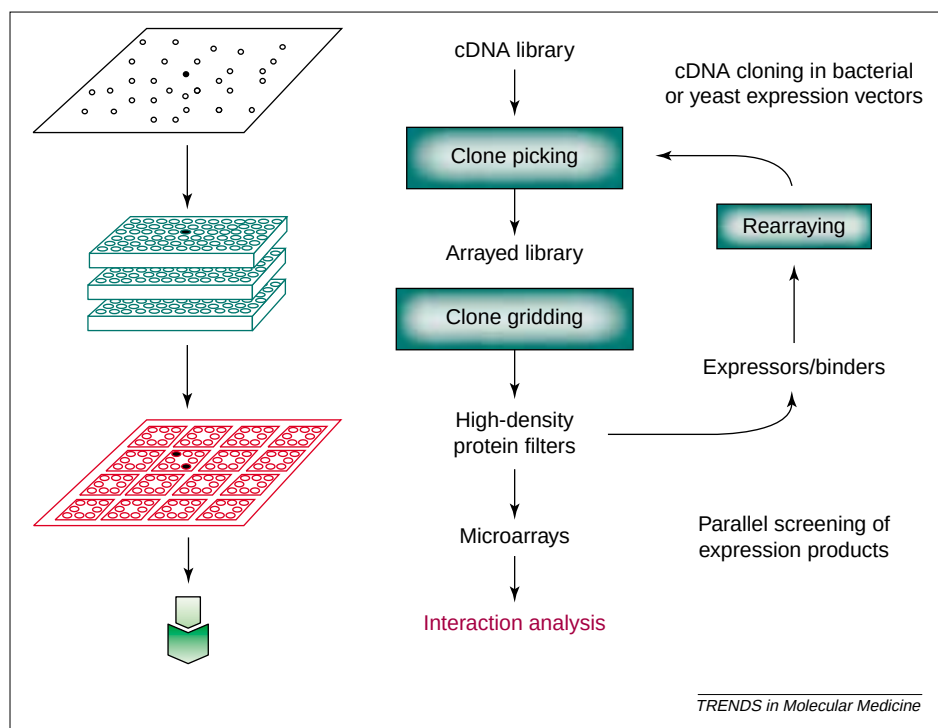


Fig. 1. High-throughput expression and analysis of unknown proteins. cDNA expression libraries are cloned in suitable expression vectors and arrayed at high density (protein filters or microarrays), using automated technology for clone picking and gridding. The arrays are screened for protein expression or ligand binding by incubation with complex molecular mixtures, and protein-expressing or protein-binding clones are subsequently rearranged into dedicated subset libraries [10,11].

expression and direct immobilization onto microtitre plates has been described as 'protein *in situ* array' (PISA) method [14]. Multidimensional chromatography of cell and tissue lysates or cellular subfractions could become an important alternative to recombinant protein expression. By consecutive chromatography and isoelectric focusing, cellular protein extracts have been directly fractionated, arrayed and detected with antibodies [15].

### Protein array technology

Array technology enables efficient high-throughput analysis of large numbers of samples [2]. Proteins can be arrayed either on flat solid phases or in capillary systems (microfluidic arrays [16]). Preferred solid phases are modified glass or filter membranes, because of their low-fluorescence background. Nitrocellulose [15] and polyvinylidene difluoride membranes [10] as well as polyamine [15,17] and polyaldehyde [9,15,18] derivatized slides have been used. Binding can be covalent (via chemical linkers) or non-covalent (via hydrophobic, van der Waals, ionic or other interactions).

Solid- and fluid-phase parameters, such as charge, viscosity, membrane pore size, pH, binding capacity and nonspecific binding, play an essential role for the generation of protein arrays. The main challenge is to preserve proteins in their biologically active shape and form. This is facilitated by attaching them in an oriented mode, using special parts (e.g. histidine tags) for binding to the (e.g. nickel-coated) solid phase, leaving active sites free and exposed for interaction [9,14].

Protein arrays are commonly printed (gridded, spotted) and imaged using the same commercially available (automated) arrayers and scanners as for DNA arrays [9,10,15,17–19]. Most arraying devices are pin-based systems, transferring nanoliter amounts of liquid either on the outside of solid pins or inside a split- or ring-shaped reservoir. Solid pins are less sensitive to variation of sample viscosity, which is a major concern of split pins and alternative micro-dispensing systems, such as ink-jet printing [20] or electrospray deposition [21]. Most imaging systems for protein arrays are fluorescence-based [9,10,15,17–19], using either direct labelling of ligands or indirect detection via labelled antibodies. Different fluorophores allow multiplexing and differential protein expression profiling. Surface-confined fluorescence detection, such as the fluorescence planar wave guide (FPWG) technique, enables detection of extremely low levels of analyte [22]. Using mass spectrometry, proteins and their

post-translational modifications can be identified in complex protein mixtures [23].

### Antibody, aptamer, PROfusion™, nano- and microfluidic arrays

The concept of antibody arrays is derived from two sources: classical immuno-capture assays and cDNA arrays. First, high-throughput probing of hundreds of compounds in complex mixtures (e.g. body fluids or tissues) can be miniaturized. Second, robust, hybridization-like antibody–antigen binding can be used in a standard way to identify all or certain marker proteins expressed from a genome [17]. Their common molecular features make antibodies ideal candidates for standard assay formats and the creation of protein-expression profiles, complementing and extending diagnostic options offered by DNA microarrays. As for other protein arrays (see above), antibodies have been arrayed as bacterial clones [24] or purified and immobilized on solid phases [17,25]. Imaging has been based on either direct fluorescent labelling of antigens [17] or sandwich assays using secondary antibodies [25] or specific antibody-binding reagents (e.g. protein A or protein L [24]). Antigen-binding fragments, such as Fab or scFv, provide simpler antibody formats, which can be affinity-selected *in vitro* using phage display and bacterial clone arrays [10,24]. Camel antibody fragments, consisting of a single immunoglobulin domain, are very robust, highly soluble in aqueous solution and offer enhanced stability as possible future array probes [26].

Based on well-established DNA microarrays, single-stranded oligonucleotides, whose specificity is optimized by *in vitro* evolution (aptamers), appear as promising new array probes [27]. Aptamers allow their covalent attachment to cognate proteins by photo-crosslinking, dramatically reducing background from other proteins in body fluids. For detection of protein binding on arrays, nonspecific protein stains can be used. PROfusion™ (Phylos, Inc., Lexington, MA, USA) arrays offer the particular advantage of self-assembly by hybridization to surface-bound DNA capture probes of a protein (phenotype) via its encoding mRNA (genotype), which are covalently linked and can be optimized by *in vitro* evolution [19]. Protein nanoarrays with 100–350-nanometer features were constructed using Dip-pen nanolithography [28]. These nanoarrays

exhibit almost no detectable nonspecific protein binding and can be screened easily by atomic-force microscopy. On the way to completely automated array systems, microfluidic arrays (Laboratory-on-a-Chip) offer capillary-based advantages (Table 1).

### Diagnostic prospects of protein arrays

The recently completed human genome sequence will provide large numbers of drug targets and cause a paradigm shift for medicine and the pharmaceutical industry. The world of diagnostics and therapeutics is changing towards more complexity, as the era of the 'blockbuster drug' will soon be over, making way for the 'package size one' treatment regime. This requires screening and cataloguing of patients' gene and protein profiles prior, during and following treatment, using efficient high-throughput technology (Table 1). Based on the concept of the arrayed library (Fig. 1), cDNA clone arrays are used for high-throughput expression and analysis of unknown proteins [2]. As a variant of this principle, antibodies were screened and selected from phage display libraries using bacterial clone arrays [24]. Building upon established DNA microarray technology, arrays of purified proteins represent the next level diagnostic tools, allowing direct analysis of gene functions [17,18]. Protein arrays enable miniaturized and highly parallel ligand binding assays for protein identification, quantification and affinity studies [1]. Protein affinity assays are used to analyze interactions between proteins such as antibodies, receptors or enzymes with other proteins, peptides, low-molecular-weight compounds, oligosaccharides or DNA. Protein profiles and interactions can already be correlated to states of cellular activity and disease. The ProteinChip/SELDI (surface-enhanced laser desorption/ionization) mass-spectrometry technology (Ciphergen, Fremont, CA, USA) is still low-throughput, but has been used for differential profiling and protein marker discovery in the realm of infectious diseases and cancer [23]. Protein expression profiles were studied in patient-matched normal and cancerous (colon, prostate, head and neck) tissues and body fluids. Over 30 proteins were detected as being either overexpressed or underexpressed in specimens from prostate cancer patients. Using new bioinformatics software (genetic algorithms and cluster analysis), proteomic patterns in serum were shown to completely discriminate cancer

from non-cancer within the ovary [29]. These findings indicate that proteomic pattern analysis might ultimately be applied as a screening tool for cancer in high-risk and general populations. This also applies to autoimmune diseases, by screening sera of patients or high-risk individuals for the presence of specific autoantibodies, using arrays of large numbers of recombinant proteins of known identity. Such arrays overcome the problems associated with variation of protein levels in conventional tissue extracts and hence improve reproducibility as a prerequisite for diagnostic use. So far, a microarray-based immunoassay involving 18 different autoantigens [5] and high-throughput arrays featuring 196 distinct biomolecules and containing the major autoantigens of eight distinct human autoimmune diseases, including systemic lupus erythematosus and rheumatoid arthritis [30], have been reported. These are the first, but extremely promising, steps towards high-throughput protein arrays as diagnostic tools, eventually arriving at the doctor's office and as over-the-counter devices.

### References

- Stoll, D. *et al.* (2002) Protein microarray technology. *Front Biosci.* C13–32
- Walter, G. *et al.* (2000) Protein arrays for gene expression and molecular interaction screening. *Curr. Opin. Microbiol.* 3, 298–302
- Nose, M. and Wigzell, H. (1983) Biological significance of carbohydrate chains on monoclonal antibodies. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6632–6636
- Moran, M.F. *et al.* (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8622–8626
- Joos, T.O. *et al.* (2000) A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* 21, 2641–2650
- Cornelis, P. (2000) Expressing genes in different *Escherichia coli* compartments. *Curr. Opin. Biotechnol.* 11, 450–454
- Lueking, A. *et al.* (2000) A system for dual protein expression in *Pichia pastoris* and *Escherichia coli*. *Protein Expr. Purif.* 20, 372–378
- Martzen, M.R. *et al.* (1999) A biochemical genomics approach for identifying genes by the activity of their products. *Science* 286, 1153–1155
- Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105
- Büssow, K. *et al.* (1998) A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* 26, 5007–5008
- Büssow, K. *et al.* (2000) A human cDNA library for high-throughput protein expression screening. *Genomics* 65, 1–8
- Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411, 107–110
- Heyman, J.A. *et al.* (1999) Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* 9, 383–392
- He, M. and Taussig, M.J. (2001) Single step generation of protein arrays from DNA by cell-free expression and *in situ* immobilisation (PISA method). *Nucleic Acids Res.* 29, E73–3
- Madoz, G.J. *et al.* (2001) Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteomics* 1, 1279–1287
- Jain, K.K. (2001) Cambridge Healthtech Institute's Third Annual Conference on Lab-on-a-Chip and Microarrays. 22–24 January 2001, Zurich, Switzerland. *Pharmacogenomics* 2, 73–77
- Haab, B.B. *et al.* (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2, 0004.1-0004.13
- MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763
- Weng, S. *et al.* (2002) Generating addressable protein microarrays with PROfusion™ covalent mRNA-protein fusion technology. *Proteomics* 2, 48–57
- Roda, A. *et al.* (2000) Protein microdeposition using a conventional ink-jet printer. *Biotechniques* 28, 492–496
- Avseenko, N.V. *et al.* (2001) Immobilization of proteins in immunochemical microarrays fabricated by electrospray deposition. *Anal. Chem.* 73, 6047–6052
- Weinberger, S.R. *et al.* (2000) Recent trends in protein biochip technology. *Pharmacogenomics* 1, 395–416
- Fung, E.T. *et al.* (2001) Protein biochips for differential profiling. *Curr. Opin. Biotechnol.* 12, 65–69
- de Wildt, R.M. *et al.* (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol.* 18, 989–994
- Huang, R.P. (2001) Detection of multiple proteins in an antibody-based protein microarray system. *J. Immunol. Methods* 255, 1–13
- Muyldermans, S. (2001) Single domain camel antibodies: current status. *J. Biotechnol.* 74, 277–302
- Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.* 74, 5–13
- Lee, K.B. *et al.* (2002) Protein nanoarrays generated by dip-pen nanolithography. *Science* 295, 1702–1705
- Petricoin, E.F. *et al.* (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577
- Robinson, W.H. *et al.* (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* 8, 295–301

### Gerald Walter\*

Biorchard Ltd., c/o Cornupia Capital Ltd.,  
27 Old Gloucester Street, London,  
UK WC1N 3XX.

\*e-mail: gerald@biorchard.com

### Konrad Büssow

#### Jörn Glökler

Max Planck Institute of Molecular Genetics,  
Innestr. 73, DE-14195 Berlin, Germany.

### Angelika Lueking

Scienion AG, Volmerstr. 7B, DE-12489 Berlin,  
Germany.