The impact of protein biochips and microarrays on the drug development process

Christoph Huels, Stefan Muellner, Helmut E. Meyer and Dolores J. Cahill

With the genome sequences of several organisms now in public databases, the scientific community has realized that it is time to prepare for the next step: the understanding of biological systems or systems biology. Whereas genes contain the information for life, the encoded proteins and RNAs fulfill nearly all the functions, from replication to regulation. At present, there is a perceived demand for high-throughput and parallel analytical devices as research tools in systems biology, and, in addition, for new concepts to extract knowledge and value from these data. Protein biochips will play a decisive role in meeting this need in the future.

Stefan Muellner
fundamenta CAPITAL
Bergische Landstr. 67
D-51375 Leverkusen
Germany
Dolores J. Cahill
Max-Planck-Institute for Molecular Genetics
Ihnestr. 73, D-14195 Berlin
Germany; and Protagen.
***Christoph Huels**
Helmut E. Meyer
Protagen
Emil-Figge-Str. 76 a
D-44227 Dortmund
Germany
*tel: +49 231 9742 890
fax: +49 231 9742 891
e-mail: christoph.huels@protagen.de

▼ High-density DNA biochips have provided researchers with a new and relevant tool. The opportunity to test 10,000 genes in parallel has increased the throughput of experiments and led to a new dimension in data handling. Indeed, the data flood was so immense that totally new bioinformatic tools, such as microarray analyzing software, had to be established. Bioinformatics has gained importance as it is required to extract valuable information from these datasets.

Two of the main reasons why DNA biochips have not reached their full potential and are unlikely to become a commodity, are poor quality and reproducibility [1]. Nevertheless, the forecasts from different sources for the development of biochip markets are promising. The overall biochip market is predicted to reach sales of approx. US$3 billion in 2004 [2]. In the past, demand for DNA biochips was driven by the need to put more targets in the R&D pipeline. The genomics era has satisfied this need by enhancing the speed of target identification and, in parallel, by increasing the amount of targets identified sixfold [3]. At the same time, the bottleneck of getting enough targets into the pipeline has been shifted to the next step – the validation of these targets.

Providing tools for target validation is, therefore, one of the major tasks of the biotechnology industry. And it seems obvious that the validation has to be made with the biomolecules that perform the function in the cell; that is, the proteins. Owing to their universal function in nature, proteins differ not only in their amino acid sequence, but also in their shape and size. Some proteins can be highly water soluble (e.g. serum and milk proteins) or virtually insoluble in any solvent (e.g. keratin and other structural proteins). In addition, the structure, function and respective stability of proteins, whether inside or outside a biological system, varies enormously. On the one hand, such individuality and diversity enables adaptation of any life form to its environment; on the other hand, it is a real, and hitherto unsolved, challenge for the designers, developers and producers of functional protein biochips.

All studies on functional analysis have to deal with these molecules. New technologies and established technology fields, such as biochemistry, genomics and proteomics, have to be applied to enhance the throughput of target candidates through the validation process.

Protein heterogeneity and technical implications

Protein biochips have the potential to speed up target validation. However, owing to the individual features of proteins, the first generation of

1359-6446/02/$ – see front matter PII: S1359-6446(02)02389-9

Table 1. Comparison of DNA and protein biochips

Properties	DNA	Protein
Structure	Uniformly built from four nucleotides	Different structures because built from 20 amino acids that can be post-translationally modified
Functionality	Functional in denatured state	3D structure is required for function
Storing	Can be stored dry	Denatured proteins can be stored dry; functional protein arrays need to be stored in 'natural' conditions
Interaction force	Hybridization (binding)	Binding
Possible interactions	DNA–DNA	Protein–protein
	DNA–RNA	Protein–peptide
	(DNA–protein)	Protein–binder (e.g. antibody)
	(RNA–protein)	Protein–small-molecule
		Receptor–ligand
		Protein–DNA
		Protein–RNA
Production	Amplification via PCR	Protein expression in different prokaryotic and eukaryotic expression systems
	Production in *E. coli*	
Solubility	Soluble in aqueous solutions	Variable solubility characteristics

protein biochips will deal with antigen-presenting protein arrays. Multiparallel ELISA (enzyme-linked immunosorbent assay)-type assays, using more or less denatured proteins, will also be developed with the aid of well-studied antibody pairs. Antigen-capturing protein arrays using antibodies or antibody-like binder proteins, which do not differ in stability and functionality at a given pH, temperature, ion strength, surface and solvent, will also be in this first generation.

One of the main issues to be addressed is the immobilization of proteins on the chip surfaces. One approach focuses on genetic engineering of the proteins for surface attachment. It might also be advantageous to immobilize the proteins in an orientated fashion to ensure that the same structural epitope, for example that covering the biologically active region of the protein, is presented for interaction.

Recombinant proteins offer the most convenient way to introduce tags into the protein at the N- or C-terminal ends. His-tags can be used to immobilize proteins using Ni^{2+}-chelating material on the chip surface [4]. Another method involves the N-terminal incorporation of serine or threonine residues into the protein sequence to enable site-specific polyethylene glycol attachment [5]. Other tags such as biotin-binding protein have also been successfully used for immobilization.

For non-recombinant proteins, other strategies for orientated immobilization have been used. These methods take advantage of existing functional groups in the protein. Most of the methods have been developed for monoclonal antibodies [6], and are designed to ensure that the antigen-binding sites of the antibody are facing away from the chip surface. The results obtained with these antibody chips have a higher and more homogeneous level of activity as compared with non-oriented immobilization [7].

The hydrophobicity of membrane-spanning or membrane-associated proteins is a major technical hurdle for their use in microarray formats. Recently, γ-aminopropylsilane (GAPS)-modified surfaces have been successfully used to immobilize vesicular solutions of membranes containing G-protein-coupled receptors [8]. Specific binding of fluorescently labeled neurotensin could be observed, indicating the proof-of-principle for this type of application.

Current status of protein chip development

Today, the generation of protein biochips is more costly and labor-intensive than that of DNA biochips. The rational to study the changes in abundance and structure of proteins is that the changes associated with proteins correlate with disease. Although analysis of the abundance of the mRNA encoding the protein of interest is the current approach, this resembles the situation of looking for a lost key near a street light even though it has been lost in the dark miles away. Although the study of defined genes, such as the HER2/neu marker in breast cancer, has led to good correlation, the study of a variety of mRNA species and their corresponding proteins has led to no correlation, to date [9,10]. This reflects not only the fact that translation of the mRNA into protein at the ribosome is regulated, but also that protein and mRNA metabolism is dynamic.

During their biosynthesis, many proteins undergo multiple modifications, such as cleavage (e.g. of leader sequences), acetylation, prenylation and phosphorylation, as well as derivatization with sugars and/or lipids. It also has to be taken into account that precursor proteins are often formed (e.g. pre-pro-insulin), and that these have to undergo enzymatic cleavage to gain their final function. All of these modifications are important for the properties of proteins, such as stability and activity. Moreover, proteins seem to be processed in different ways, leading to different functionality in the course of a disease; this has been shown in Alzheimer's disease [11]. Some proteins are regulated tightly in a tissue, or even at the single-cell level as in cystic fibrosis [12]; diseased states such as these cannot be studied by simply looking at mRNA expression.

Nearly all targets currently used for drug development are proteins. The ultimate goal for validating the vast number of protein targets is, therefore, to gain information about their relevance in a diseased or healthy organism, and about their differential expression and function in specific organs, cells and tissues. However, the information obtained by such means is not sufficient to accelerate preclinical and clinical studies. Furthermore, it is important that the data generated will serve as a basis to design a new relevant pharmacological animal model that shows a clear link to human disease and can be accepted by clinical pharmacologists. At present, one of the most promising approaches for performing such studies is the use of protein microarrays or protein biochips. As mentioned previously, DNA and proteins are very different molecules; Table 1 gives an overview of some of their major differences.

Protein arrays generally comprise recombinant proteins that are robotically spotted onto small, flat surfaces. Their production can be highly automated, using either pin-based or micro-dispensing liquid-handling robots. The source of proteins for low-density arrays is generally a subset of the proteins or antibodies of interest [13]. Recent developments include the generation of low-density protein arrays on filter membranes, such as the universal protein array system (UPA), which is based on the 96-well microtitre plate format [14], enabling the study of protein–protein and protein–ligand interactions, as well as of protein affinities on filters.

Other methods of generating protein microarrays is by printing the proteins onto an optically flat glass plate containing 96 wells, formed by an enclosing hydrophobic Teflon mask [15]. Standard ELISA techniques and a scanning charge-coupled device (CCD) detector have been used for imaging arrayed antigens. Other approaches to generate protein microarrays have been reported using either photolithography of silane monolayers [16] or gold [17], combining micro-wells with micro-sphere sensors [18] or with ink-jetting onto a polystyrene film [19].

Automation technologies

For the generation of high-density protein arrays, automation technologies have been developed for the production of cDNA expression libraries, high-throughput protein expression and large-scale protein analysis [20,21]. The first method involved robotic technology to array bacterial colonies from a human fetal brain cDNA expression library (hEx1) onto a polyvinylidene difluoride membrane to produce high-density colony filters. *In situ* expression of recombinant fusion proteins was induced and detected using an antibody against a His-tag-containing epitope. After washing, the filters were then incubated with the appropriate labeled secondary antibody and substrate. The same clones were positive in hybridizations with the cDNA probe encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Fig. 1).

This approach has been extended by automated spotting and immobilization of purified protein microarrays from liquid expression cultures [20]. Sharp and well-localized signals

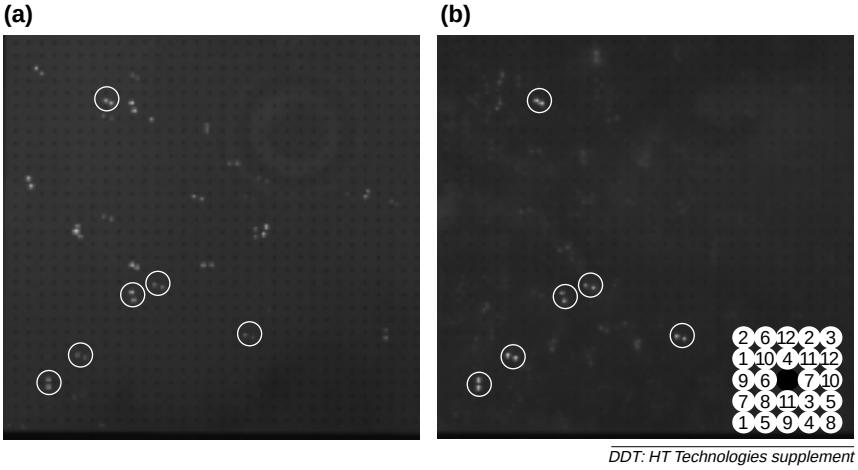

Figure 1. Detection of the DNA and protein levels of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA clones from the same gene library. Clones containing parts of the GAPDH cDNA do not necessarily express the protein. *Escherichia coli* colonies were grown on filters, lysed and tested for (a) the presence of the respective insert using DNA-hybridization with a GAPDH cDNA probe; or (b) their ability to express GAPDH by screening for binding with a polyclonal anti-GAPDH antibody [20]. Clones showing signals in both tests are circled [5 × 5 gridding pattern is shown in the lower right corner of (b)].

enabled the detection of 250 attomole or 10 pg of a spotted test protein (GAPDH) by a specific antibody, which showed that such protein microarrays provide the means for sensitive gene expression and antibody specificity screening using minimal amounts of reagents.

Similar studies have been performed with 5800 different open reading frames in yeast. The corresponding proteins were expressed in yeast and purified before spotting onto glass slides. This protein biochip was suitable for identifying novel calmodulin- and phospholipid-interacting proteins [22].

Protein biochip technologies include protein identification, quantification and affinity studies. Additionally, technologies for the detection of proteins, such as mass spectrometry (MS) or biomolecular interaction analysis using surface plasmon resonance (BIAcore) technology, can be included under the broad term of protein biochip technologies [23]. These technologies have contributed greatly to interaction studies in a chip format. For capture and analysis of specifically labeled proteins, ligand-coated surfaces are available, such as the SELDI ProteinChip® (surface-enhanced laser desorption–ionization) distributed by Ciphergen Biosystems (http://www.ciphergen.com), and various BIAcore chips (Biacore; http://www.biacore.com). SELDI and BIA–MS technology combine affinity capture and mass analysis into a novel and efficient means for the characterization of proteins and their interaction partners, and have applications such as identification of novel biomarker candidates, epitope mapping and receptor–ligand or protein–drug interaction studies [24].

Having access to many different recombinant proteins, which are expressed and purified, is a prerequisite for the generation of well-defined high-content protein biochips. The source of these proteins, methods for protein production and purification, quality control of the proteins, and the storing and spotting of the proteins, are major challenges for the near future. cDNA libraries might enable the efficient and fast cloning of such genes and facilitate the subsequent automated expression of the encoded proteins.

Application to the R&D pipeline

Proteins as targets are dominating pharmaceutical R&D. Ligand–receptor interactions and enzymes are the vast majority, comprising ~45% and 28%, respectively, of the targets [25]. Additionally, many therapeutic proteins, especially humanized antibodies, are in clinical development. The ultimate tool for high-throughput and significant screening would be to test new leads or new targets in a highly parallel manner. Some examples for applications in this direction already exist. Recently, an immunosensor array has been developed that enables the simultaneous detection of clinical analytes [26]. Here, capture antibodies and analytes were arrayed onto microscope slides using flow chambers in a cross-wise fashion. This current format is low-density (6×6 pattern) but has high-throughput potential as it involves automated image analysis and microfluidics; it is already becoming one of the future formats for enzyme activity testing and other assays [27]. In another study, small sets of active enzymes were immobilized in a hydrophilic gel matrix. Enzymatic cleavage of the substrate could be detected and inhibitors blocked the reaction [28]. More recently, an enzyme array that is suitable for assays of enzyme inhibition has been reported [29].

Initial publications in the area of receptor–ligand interaction studies in a microarray format have shown that the interaction of immobilized compounds and proteins in solutions can be determined [22,30,31]. This technology enables HTS of ligand–receptor interactions with small sample volumes. The proof-of-principle to study differential protein expression using such microarrays has also been shown [32]. A robotic device was used to print hundreds of specific antibody or antigen solutions in an array onto the surface of derivatized microscope slides. Specific antibody–antigen interactions could be quantitatively measured using a reference standard. These results suggest that protein microarrays can provide a practical means to characterize patterns of variation in hundreds-of-thousands of different proteins in clinical or research applications.

Another recent application of protein arrays is their use in identifying highly specific antibody–antigen interactions by screening protein array filters with antibody fragments. This novel approach enables the generation of specific antibodies without requiring the immunization of animals or *in vitro* phage- or ribosome-display technologies [33]. This approach has been further developed by Cahill and colleagues to screen, in parallel, ~2500 different human recombinant purified proteins on a single microscope glass slide for recognition by auto-antibodies from sera of autoimmune patients (D. J. Cahill, unpublished). The approach led to the detection of known disease-specific autoimmune markers, common markers detected in a variety of autoimmune diseases, and novel markers that are characteristic for the respective autoimmune disease. Similar studies have been done with a subset of proteins representing 196 major auto-antigens on the biochip surface with sera from autoimmune patients [34]. Reproducible disease-specific patterns of antigen recognition could be observed.

Outlook and comparison to conventional proteomics technologies

The work reviewed here opens the vision for future applications (Fig. 2). The ultimate value of the recent combination of efforts in genomics, proteomics and biochip technologies, and their impact on the overall drug development process, are striking. For the first time, tools are available to study disturbances within a biological system on the gene sequence and expression

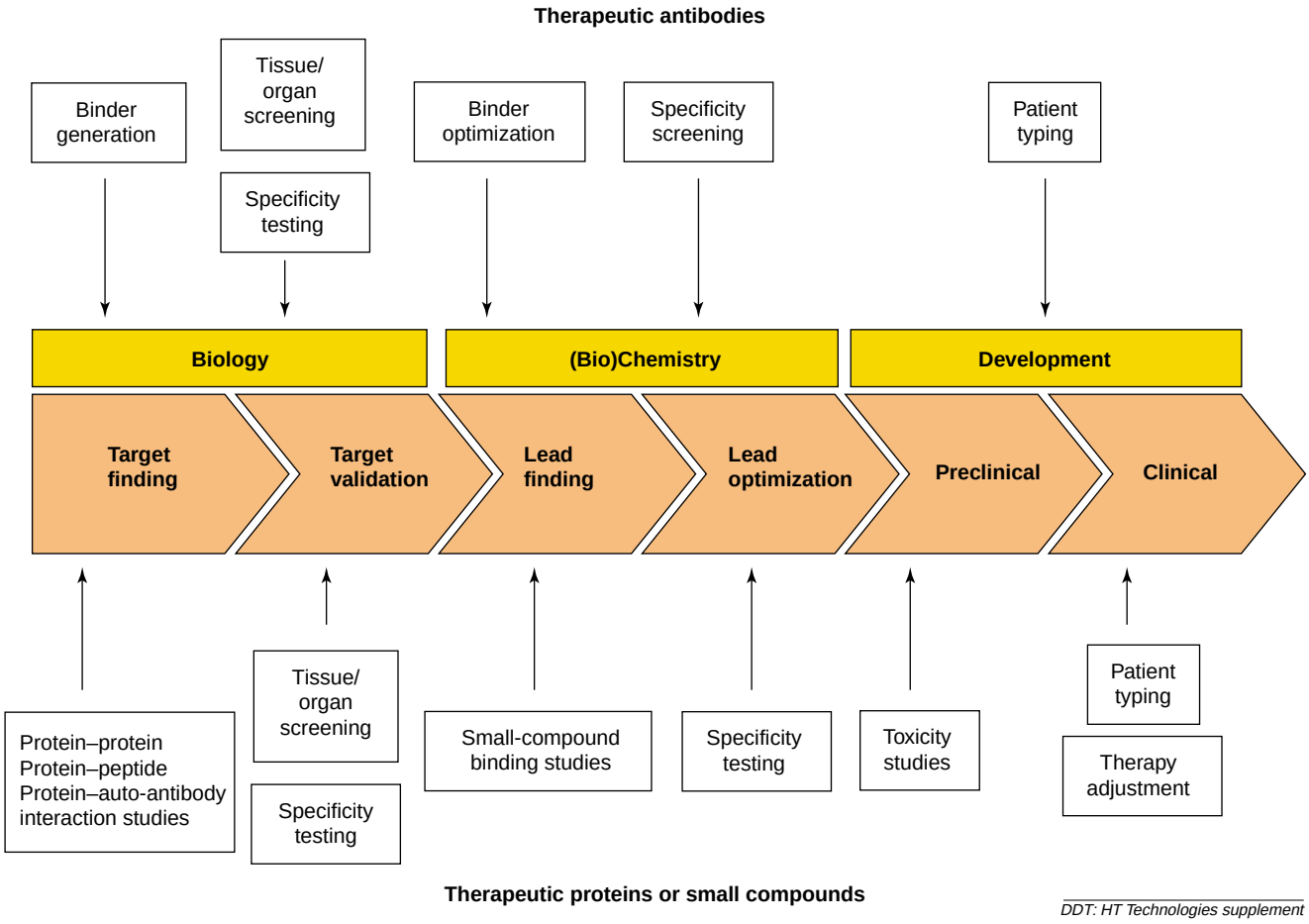

Figure 2. Protein biochips and their possible contributions to pharmaceutical R&D. Currently, the main applications of protein biochips are in the study of protein–protein interactions and the interaction between small compounds and protein immobilized onto the biochip surface. These studies could lead to new approaches for target identification and validation and for lead-compound detection and optimization. Clinical phases can also potentially profit from the contributions of protein biochips. Patient-typing and -stratification could be performed in this new assay format in a multi-parallel way. In addition, the development of therapeutic monoclonal antibodies (mABs) can profit from protein biochip dedvelopments. Testing the specificity of a new mAB against thousands of different proteins in parallel with minimum amounts of sample has the potential to generate more-specific therapeutic mABs in a shorter timeframe.

level. In addition, protein biochips provide data about protein expression that correlate directly with the function or dysfunction of a given system. Applications such as target identification and characterization, target validation, diagnostic marker identification and validation, preclinical and clinical study monitoring, and patient typing (i.e. selection of the right patient for the best drug therapy), seem to be feasible. The multi-parallel diagnostic possibilities of protein biochip applications have the potential not only to enable the optimization of preclinical, toxicological and clinical studies through better selection and stratification of individuals, but also to affect how diagnostics are used in drug development.

Classical proteomics approaches for expression proteomics (e.g. 2D-gel electrophoresis) will continue to contribute to the target identification and validation steps. In contrast to protein biochips, these technologies lack the throughput and capacity to analyze multiple probes in parallel [35–37]. The cell-map proteomics approaches, such as yeast two-hybrid systems, have the advantage of easy read-out and high-throughput. These studies have also proven to be applicable to virtually genome-wide studies [35,38]. For protein–compound interaction studies, the compound still has to interact inside the nucleus with the target molecules, and the proteins of interest have to be expressible as fusion proteins in yeast. Protein biochips offer the potential to perform these studies in micro-wells in solution. In addition, the proteins can be expressed in a variety of production systems, including mammalian cells, to ensure proper folding and modification for native expression.

Potential applications for protein biochips in pharmaceutical R&D are widespread (Fig. 2). Miniaturized ELISAs will be feasible,

as well as protein arrays for antibody screening and characterization. Next-generation applications will then focus on protein–protein, protein–peptide and protein–small compound interaction assays to serve a faster pharmaceutical R&D process. One of the major challenges is the supply of these assays with well-characterized proteins and their respective binders (e.g. monoclonal antibodies) for the development of these assay formats.

References

1 Calkins, K. (2001) Holding all the chips. *BioCentury The Bernstein Report on Biobusiness*, 9, A1–A3

2 Frost & Sullivan (2001) The World Biochips Market, Report #7761–55

3 The Boston Consulting Group (2001) *A revolution in R&D,The Impact of Genomics*, pp. 1–15 (http://www.bcg.de/start.asp)

4 Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105

5 Gaertner, H.F. *et al.* (1996) Site-specific attachment of functionalized poly (ethylene glycol) to the amino terminus of proteins. *Bioconjug. Chem.* 7, 38–44

6 Lu, B. *et al.* (1996) Oriented immobilization of antibodies and its applications to immunoassays and immunosensors. *Analyst* 121, 29R–32R

7 Vijayendran, R.A. and Leckband, D.E. (2001) A quantitative assessment of heterogeneity for surface-immobilized proteins. *Anal. Chem.* 73, 471–480

8 Fang, Y. (2002) Membrane protein microarrays. *JACS commun.* 124, 2394–2395

9 Lohrisch, C. and Piccart, M. (2001) HER2/neu as a predictive factor in breast cancer. *Clin. Breast Cancer* 2, 129–135

10 Anderson, L. and Seilhammer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537

11 Masters, C.L. and Beyreuther, K. (1998) Alzheimer´s disease. *Br. Med. J.* 316, 446–448

12 Jiang, Q. and Engelhardt, J.F. (1998) Cellular heterogeneity of CFTR expression and function in the lung: implications for gene therapy of cystic fybrosis. *Eur. J. Hum. Genet.* 6, 12–31

13 Joos, T.O. *et al.* (2001) A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* 21, 2641–2650

14 Ge, H. (2000) 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 (i–vii)

15 Mendoza, L.G. *et al.* (1999) High throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biotechniques* 27, 778–788

16 Mooney, J.F. *et al.* (1996). Patterning of functional antibodies and other proteins by photolithography of silane monolayers. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12287–12291

17 Jones, V.W. *et al.* (1998). Microminiaturized immunoassays using atomic force microscopy and compositionally patterned antigen arrays. *Anal. Chem.* 70, 1233–1241

18 Michael, K.L. *et al.* (1998) Randomly ordered addressable high-density optical sensor arrays. *Anal. Chem.* 70, 1242–1248

19 Ekins, R.P. (1998) Ligand assays: from electrophoresis to miniaturized microarrays. *Clin. Chem.* 44, 2015–2030

20 Buessow, 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

21 Lueking, A. *et al.* (1999) Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* 270, 103–111

22 Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105

23 Natsume T. *et al.* (2001) BIA–MS–MS: biomolecular interaction analysis for functional proteomics. *Trends Biotechnol.* 19, 28–33

24 Weinberger, S.R. *et al.* (2000) Recent trends in protein biochip technology. *Pharmacogenomics* 1, 395–402

25 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964

26 Rowe, C.A. *et al.* (1999) An array immunosensor for simultaneous detection of clinical analytes. *Anal. Chem.* 71, 433–439

27 Cohen, C.B. *et al.* (1999) A microchip-based enzyme assay for protein kinase A. *Anal. Biochem.* 273, 89–97

28 Arenkov, P. *et al.* (2000) Protein microchips. Use for immunoassays and enzymatic reactions. *Anal. Biochem.* 278, 123–131

29 Park, C.B. and Clark, D.S. (2002) Sol-Gel encapsulated enzyme arrays for high-throughput screening of biocatalytic activity. *Biotechnol. Bioeng.* 78, 229–235

30 Mangold, U. *et al.* (1999) Identification and characterization of potential new therapeutic targets in inflammatory and autoimmune diseases. *Eur. J. Biochem.* 266, 1184–1191

31 MacBeath, G. *et al.* (1999) Printing small molecules as microarrays and detecting protein–ligand interactions en masse. *J. Am Chem. Soc.* 121, 7967–7968

32 Haab, B.B. *et al.* (2001) Protein microarrays for highly parallel detection and quantification of specific proteins and antibodies in complex solutions. *Genome Biol.* 2, 4.1–4.13

33 Holt, L.J. *et al.* (2000) By-passing selection: direct screening for antibody–antigen interactions using protein arrays. *Nucleic Acids Res.* 28, e72 (i–v)

34 Robinson, H.R. *et al.* (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* 8, 295–301

35 Blackstock, W.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121–127

36 Marshall, T. *et al.* (2002) Proteomics and its impact upon biomedical science. *Br. J. Biomed. Sci.* 59, 47–64

37 Muellner, S. The impact of proteomics in industrial processes and products. In *Proteomics of microorganisms: Fundamental aspects and application* (Hecker, M. and Mueller, S., eds), Springer Verlag (in press)

38 Gavin, A-C. *et al.* (2002) Functional characterization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147