general basis. In our opinion, the variety of genomic questions has to be matched by the flexibility of the technology. Especially in the microarray field, flexibility and accessibility will be the key for successfully unravelling the information produced by the various sequencing projects. Although, at present, quite a variety of genomic questions are addressed by microarray technology, one can foresee that, besides gene expression profiling, resequencing and SNP typing, new genomic applications will become popular in the near future. To be able to address these applications using microarrays, the technology itself has to be maximally flexible. As sequence information for an increasing number of organisms becomes available in public databases, decentralized facilities will become more and more important, because not technologists, but biologists, will be the ones to set up the genomic assays. No centralized facility will be able to have the huge core know-how for the many genomic assays that Nature offers for microarray applications. This kind of scenario will allow even exotic or commercially unpopular organisms to be assayed using microarray technology. Therefore we believe that giving researchers themselves the possibility to program the individual genomic assays at their laboratories will speed up the acquisition of genomic know-how dramatically. This tendency will be even more pronounced through the use of the Internet as an exchange medium, whereby genomic assays will be sent as digital files between collaborating laboratories, or can easily be downloaded like other conventional software.

References
1. http://ergo.integratedgenomics.com/GOLD (Genome Online Database)

Received 19 December 2001

The ProteinChip® Biomarker System from Ciphergen Biosystems: a novel proteomics platform for rapid biomarker discovery and validation
K. Chapman
Ciphergen Biosystems Ltd, Surrey Technology Centre, The Surrey Research Park, Guildford, Surrey GU2 7YG, U.K.

Abstract
Protein biochips are emerging in two distinct formats. The first involves high-density immobilized arrays of recognition molecules (e.g. antibodies), where target binding is monitored indirectly (e.g. via fluorescence). This technology is in its infancy, being limited by the availability of suitable binding molecules that can cope effectively with protein diversity. The second format involves the capture of proteins by biochemical or intermolecular interaction, coupled with direct detection by MS. This technology is available as the ProteinChip® Biomarker System. ProteinChip technology uses surface-enhanced laser desorption/ionization processes to analyse proteins directly from biological samples. Chromatographic surfaces are placed on to ProteinChip Arrays and used to capture subclasses of proteins, dependent

Key words: biomarker, mass spectrometry, protein biochip, proteomics, SELDI.
Abbreviations used: 2-DE, two-dimensional protein gel electrophoresis; Qq, quadrupole–quadrupole; SELDI, surface-enhanced laser desorption/ionization; TCC, transitional cell carcinoma; TOF-MS, time-of-flight MS.
1 e-mail kchapman@ciphergen.com

© 2002 Biochemical Society 82
on their physical properties. Time-of-flight MS then assigns native molecular masses to the captured proteins. Reproducible protein profiles can be generated from crude biological fluids (e.g. cell lysates, urine or serum). The technology is being applied to a wide range of disciplines, from plant sciences to cancer research, and will be reviewed here.

**Technologies for proteomics**

In recent years there has been a vast amount of activity in genome sequencing projects and the analysis of gene (mRNA) expression. It is hoped that a better understanding of the genome, its expression patterns and functional products will lead eventually to an improved understanding, diagnosis and treatment of disease. DNA- and RNA-based technologies are extremely powerful, yet knowledge of these areas cannot be used to predict patterns of protein expression, or possible downstream processing of these proteins. Several studies have illustrated a poor correlation between gene expression and protein abundance [1]. In addition, certain body fluids (such as serum, urine and saliva) can only be described in terms of their protein composition - mRNA analysis is not applicable. These types of samples are also highly desirable for diagnostic screening, as their retrieval is considered non-invasive. Therefore there is a clear need for the direct analysis of protein expression; a research area described as proteomics.

At present, the most widely used tool for proteomic studies is two-dimensional protein gel electrophoresis (2-DE). Proteins are separated first based on their isoelectric point (pI) and then by their molecular mass. This technique has a high resolving power and can display several thousand proteins. However, the technique has a number of well recognized limitations [2]. It is generally accepted to be labour intensive, and the generation of reproducible gels requires considerable technical expertise. In addition, there are certain types of proteins that are not well represented by 2-DE, including low-molecular-mass proteins (< 20 kDa) and those with extreme pIs or hydrophobicity.

Several research groups and companies are striving towards generating protein biochips, in order to miniaturize proteomics, and the technology is emerging in two distinct formats. The first of these is an approach analogous to DNA biochips, where high-density arrays of immobi-

lized recognition molecules (e.g. antibodies) are prepared, and target binding is monitored indirectly (e.g. by fluorescence). This technology is limited by the availability of suitable binding molecules. Recognition molecules must be generated that interact with all the known (or predicted) gene products. Clearly this is not as straightforward as the generation of DNA probes used on DNA microarrays. Also, recognition molecules will profile binding epitopes rather than the functional protein molecules. An individual protein may exist in several processed forms (truncated or post-translationally modified) that may not be discriminated by the biochip. The second protein biochip format involves the capture of proteins by biochemical or intermolecular interaction followed by direct detection by MS. This technology is commercially available from Ciphergen Biosystems as the ProteinChip® Biomarker System, and will be the focus of this paper.

**Ciphergen's ProteinChip technology**

The ProteinChip System is based on proprietary surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) MS [3], which combines protein capture on chromatographic surfaces with MS. The ProteinChip Arrays are derivatized with affinity matrices that mirror the properties of conventional chromatography media such as reverse phase; anion and cation exchange; metal affinity; and normal phase (Figure 1). Depending upon the type of chromatographic surface and wash condition applied, different sets of proteins will be isolated from the crude sample (Figure 2).

The analytical process involves the application of 1 µl or more of a crude sample (e.g. tissue extract or bio-fluid) on to a spot of a ProteinChip Array. After binding, the spots are washed with buffer to remove non-specifically bound analytes. Increasing the stringency of the wash conditions (modulating the pH or salt concentrations) selects for different populations of proteins to be retained on the surface. A final wash with a weakly buffered solution can be used to remove residual salt or detergents prior to MS analysis. Energy-absorbing molecules are applied to the spot as a final step before drying. The energy-absorbing molecule is used to facilitate the desorption and ionization of retained proteins in the ProteinChip Reader, a TOF-MS instrument.

The major applications of the ProteinChip System are: (1) expression profiling or proteomics; (2) purification development and
monitoring; (3) protein characterization and identification; and (4) protein interaction assays (biochemical arrays; see Figure 1). This paper will focus on proteomics applications and, in particular, the search for clinically relevant biomarkers.

**Expression profiling and marker validation**

One of the difficulties in the use of 2-DE to discover disease biomarkers arises in the validation stage. Analysing a small sample set by 2-DE can discover potential changes in protein expression. These proteins can then be excised from the gel and attempts can be made to identify the protein via peptide mapping (discussed in more detail below). However, this newly identified biomarker has not been firmly correlated with the disease. This must be achieved by examining the expression of the potential biomarker in a much larger sample set. This is not feasible using 2-DE, and typically the assay is moved to a new platform (e.g. ELISA), which often requires antibody generation.

In contrast, protein expression profiling using the ProteinChip System can be likened to mRNA expression studies using DNA microarrays (Figure 3). In both cases, the same technology platform can be used for biomarker discovery and validation. For ProteinChip technology, a small sample set would be analysed using a range of ProteinChip Arrays and wash conditions. Once potential biomarkers are discovered, it is possible to focus on a particular surface and wash condition and rapidly screen a much larger sample set in order to validate the observed changes in protein expression.

**Figure 1**

Diagram of various ProteinChip Array surfaces

IMAC, immobilized metal affinity chromatography.

**Chemical Surfaces**

- Hydrophobic
- Ionic
- IMAC

**Biochemical Surfaces**

- Antibody
- DNA
- Enzyme
- Receptor

**Figure 2**

Rat liver lysate exposed to three different ProteinChip Array surfaces

(A) Reverse phase, water wash; (B) anionic exchange, pH 6.5 wash; (C) cationic exchange, pH 4.5 wash.

**Mass Range**

3000-17000 Da

A

B

C
There is an increasingly important requirement to move away from 2-DE to an alternative platform for marker validation.

<table>
<thead>
<tr>
<th>mRNA expression</th>
<th>Protein Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA and Oligo Arrays</td>
<td>Proteinchip® Arrays</td>
</tr>
<tr>
<td>Find mRNA with altered Expression Pattern</td>
<td>Protein Profile Changes</td>
</tr>
<tr>
<td>Confirm Changes (Marker Validation)</td>
<td>Confirm Changes (Marker Validation)</td>
</tr>
<tr>
<td>Validation on the same technology platform: Examine a smaller number of genes/proteins using a larger sample set</td>
<td>Identify Validated Marker</td>
</tr>
</tbody>
</table>

Validation assay must be moved to a new platform (ELISA etc).

expression and correlate them with the onset of disease. Once a biomarker had been shown to have good diagnostic potential, only then would identification of the protein species be attempted.

**Protein expression profiling in clinical studies**

Protein expression profiling by any method requires the reproducible visualization of a set of proteins so that samples from different treatment or disease groups can be confidently compared. The Proteinchip System offers several advantages compared with other proteomic approaches. Protein expression profiling on the ProteinChip System utilizes the chromatographic ProteinChip Arrays to select a subset of proteins based on their biophysical properties (such as pI or hydrophobicity). In this way, hundreds of proteins can be detected from a single active spot without the need for specific recognition molecules. This overcomes the first challenge encountered with high-density protein arrays described above, namely the generation of hundreds or thousands of specific capture molecules. A chromatographic ProteinChip Array can be prepared and analysed in approx. 1 h, and as each array contains eight or 16 active spots, many samples can be analysed in parallel. This represents a more rapid profiling approach than is realistically achievable with 2-DE analysis.

There are also several specific problems posed by clinical samples. Reproducibility of data is affected by sample variation as well as the robustness of the analytical technique. Sample variation is relatively easy to control in model systems. Cell culture growth conditions can be well defined, and even animal models can have controlled diet regimes. However, biological variation within clinical samples is much harder to control and can be affected by sample contamination, or sample processing and storage (transfer time from clinic to laboratory). In order to generate statistically meaningful data from variable sample sets, a large sample population must be examined; a process that is feasible with Proteinchip proteomics. Further complications arise with regard to sample heterogeneity. With prostate cancer, for example, regions of tumour are 'contaminated' with a large number of non-transformed cells, making studies of bulk, homogenized samples difficult to interpret [4]. This problem has been tackled using laser capture microdissection to isolate pure populations of cancerous cells [4–6]. Previously, protein expression studies in samples obtained by laser capture microdissection have been exceedingly
difficult, as ~50000 cells are needed for 2-DE analysis. However, as few as 2000 cells can be profiled using ProteinChip technology [4–6], illustrating the utility of the ProteinChip System when sample quantities are limiting.

Looking for disease biomarkers – bladder cancer

There are many examples of the use of the ProteinChip System to look for disease biomarkers [7–9]. One such example is the search for markers of transitional cell carcinoma (TCC) in urine [10]. There is a need to develop a non-invasive method to diagnose TCC of the bladder. Currently, the most reliable diagnosis is achieved by cystoscopic examination and bladder biopsy, which are both invasive and labour intensive. Biomarkers in urine would be highly desirable, due to the non-invasive nature of this bio-fluid. Using the ProteinChip System, 94 urine samples from patients with TCC, other urogenital diseases and healthy subjects were analysed using strong anion-exchange ProteinChip Arrays [10]. Multiple protein changes were detected reproducibly in the TCC group. The correlation of individual markers with TCC was relatively low, but combining markers (using conventional statistical methods) increased the sensitivity of the assay. This combinatorial approach gave a sensitivity of 78% in detecting grade I and II carcinomas, compared with 20–30% by voided urine cytology [11]. The ability to rapidly monitor many proteins at once is a marked advantage over individual protein assays (such as immunoassays).

Multicomponent biomarker analysis

There is an increasing need to generate software that can deconvolute large data sets in order to find common ‘themes’ that may have a diagnostic use. These analyses will tend to rely on patterns rather than individual biomarkers. One such approach is cluster analysis, which generates groupings of like samples based on similarities in the data produced from these samples. This is an unbiased analysis, as no prior information on sample phenotype is used. However, it can be difficult to determine the rationale employed by the software to group the data.

Ciphergen Biosystems can provide Biomarker Patterns™ Software, which is a classification tool. An initial training sample set, with known phenotypes, is used to stratify the data. The resulting set of stratification rules is displayed as a ‘tree’ diagram that can be used to classify unknown samples. The variables (protein masses) used to generate the classification tree are clearly displayed, and can be used to prioritize studies to identify these potential biomarkers.

Biomarker purification strategies

For diagnostic needs, a pattern of biomarkers may be deemed sufficient information to stratify disease groups. However, in order to understand disease progression better and to explore possible drug targets, it is necessary to identify the disease biomarkers as a first step towards understanding their function. As many proteins are retained simultaneously on the ProteinChip Arrays, it is necessary to enrich potential biomarkers to enable identification. Devising purification strategies is greatly assisted by the chromatographic affinity information gained during the discovery process.

A biomarker binding to a strong-anion-exchange ProteinChip Array can be enriched by utilizing a strong-anion-exchange column in the purification scheme [12]. Microspin columns are available for this purpose. The ProteinChip System then becomes the screen for the column fractions, allowing rapid monitoring of biomarker elution.

Biomarker identification – prostate cancer

Once a potential marker has been purified, many techniques can be employed to ascertain its identity (e.g. Edman degradation). However, protein quantities can be low, and often a semi-purified biomarker is run on a one-dimensional gel as the final preparative step, to prevent further sample losses. The gel band can be excised and exposed to tryptic peptide mapping, a method commonly employed to identify 2-DE gel spots. The ProteinChip System can be used to analyse the digest peptides, and the masses used to search against protein databases for possible matches to known proteins. This method for protein identification is heavily dependent upon the data available in the databases. Some organisms are not well represented, and when relying on predicted protein products from gene sequences, the ‘protein’ match may be of unknown function. In these cases, sequence information is a vital aid to finding the protein’s identity. Ciphergen have developed a ProteinChip Interface for quadropole–quadropole (Qq) TOF mass spectrometers, which are capable of performing collision-induced dissociation [3]. This system will allow researchers to sequence peptides directly from the surface of ProteinChip.
Arrays. The complementarity of the ProteinChip System and ProteinChip QqTOF was demonstrated by work to differentiate between benign prostatic hyperplasia and prostate cancer [3]. Seminal plasma samples from patients diagnosed with benign prostatic hyperplasia or prostate cancer were screened using the ProteinChip System. Several potential biomarkers for prostate cancer were discovered, and on-chip isolation was achieved for a 5.7 kDa marker. On-chip tryptic digest results suggested that the protein was semenogellin 1, but this is a much larger protein of molecular mass 52 kDa. The 5.7 kDa peptide was then sequenced using Proteinchip QqTOF-MS, revealing that the marker was seminal basic protein, a proteolytically derived fragment of semenogellin 1.

Future directions
The ability of the ProteinChip System to screen large sample sets; test simultaneously for multiple protein changes; and deconvolute the data with dedicated software shows that it has the potential to improve disease diagnosis. Continued development of the technology is focused on creating a greater diversity of capture surfaces and supporting products, such as fractionation and purification kits. In addition, as it is applied more frequently to large-scale projects, then the automation of sample preparation, array loading and reading will follow.

References
1 Anderson, L. and Seilhamer, J. (1997) Electrophoresis 18, 533–537

Received 19 November 2001