Review TRENDS in Pharmacological Sciences Vol.27 No.5 May 2006


# Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry

**Judith Y.M.N. Engwegen[1], Marie-Christine W. Gast[1], Jan H.M. Schellens[2,3] and Jos H. Beijnen[1,3]**

[1]Department of Pharmacy and Pharmacology, The Netherlands Cancer Institute and Slotervaart Hospital, Louwesweg 6, 1066 EC Amsterdam, The Netherlands
[2]Department of Medical Oncology, The Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[3]Department of Biomedical Analysis, Division of Drug Toxicology, Faculty of Pharmaceutical Sciences, Utrecht University, PO Box 80.082, 3508 TB Utrecht, The Netherlands

**Recently, the focus of cancer research has expanded from genetic information in the human genome to protein expression analyses. Because this 'proteome' reflects the state of a cell, tissue or organism more accurately, much is expected from proteomics to yield better tumour markers for disease diagnosis and therapy monitoring. Some current proteomic technologies are particularly suitable for protein profiling in the search for new biomarkers. Surface-enhanced laser desorption ionization time-of-flight mass spectrometry has been used frequently, highlighting many new proteins as biomarkers (e.g. for ovarian, breast, prostate and colorectal cancer). However, it is becoming increasingly recognized that reproducibility and validation of these biomarkers should be addressed carefully, as should their origin and identity. If these efforts are made, protein profiling can contribute to the better diagnosis of patients and the optimization of their treatment.**

## Why proteomics?

In recent years, there has been an enormous growth in the use of genome information in science. This genome information has greatly expanded the insight into the genetic basis of cancer. The advanced understanding, however, has also made it clear that gene analysis alone does not tell the whole story [1]. Owing to the alternative splicing of both mRNA and proteins, combined with protein posttranslational modifications, one gene can encode a considerable protein population. The proteome encompasses all proteins that result from the genome of cells, a tissue or an organism. It is not a static parameter, as is the genome, but a dynamic collection of proteins that reflects both the intrinsic genetic programme of the cell and the impact of its immediate environment. Compared with the genome, the proteome provides a more realistic view of a biological status and is, therefore, expected to be more useful than gene analysis for evaluating, for example, disease presence, progression and response to treatment. Thus, proteomics can bridge the gap between the genome sequence and cellular behaviour.

The potential of proteomics has been discovered in various medical areas (e.g. infectious diseases [2], Alzheimer's disease [3] and cardiovascular diseases [4]). The interest for proteomics within the oncology field is most obvious: cancer is a DNA disease that originates in mutated genes and leads to aberrant protein expression. The expectations were, therefore, that cancer proteomic studies might identify disease-related biomarkers for early cancer diagnosis and new surrogate biomarkers for therapy efficacy and toxicity, but also for guidance of optimal anticancer drug combinations, enabling tailor-made therapy. Furthermore, they could lead to new pharmacological targets. If these markers could be measured in a readily accessible body fluid such as serum, not requiring a tissue biopsy, it would have a major impact on future cancer diagnosis and treatment monitoring.

Yet studying a whole proteome for new cancer biomarkers requires systematic analysis of the many and diverse properties of proteins, their structure, function and expression in biological systems during health and disease: an immense challenge. Recent advances in analytical technologies for protein analysis such as mass spectrometry (MS) [5] and protein microarrays [6] have brought such large-scale proteomic analyses within reach. In this article, we summarize several of these technologies and their usefulness in the search for new biomarkers. We focus on surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) MS because its introduction has led to a vast increase in the number of publications about new serum biomarkers for cancer. However, it has been the subject of strong debate. We highlight several of its applications in oncology and touch upon some crucial issues.

*Corresponding author:* Engwegen, J.Y.M.N. (apjen@slz.nl).
Available online 4 April 2006

 0165-6147/$ - see front matter  doi:10.1016/j.tips.2006.03.003

## Proteomic technologies

With proteomics defined as the study of all proteins in a biological system, it is obvious that much is demanded from the analytical technologies used. More than 500 000 proteins have been estimated to comprise the human proteome, compared with ~40 000 genes in the human genome [1]. These proteins are free or membrane bound, sequestered in a specific cellular compartment or captured in a tight protein–protein interaction. The bulk of the easily accessible serum proteome consists of only a small number of abundant proteins such as albumin and immunoglobulins, which complicates the detection of the many low-abundance proteins. Thus, the major tasks are to separate and analyse the whole proteome and to process the massive amount of data into meaningful results using statistics and bioinformatics.

Clinical proteomics is aimed at finding proteins for which altered behaviour in disease (with regard to structure, function, interaction or expression) forms a starting point for improvements in clinical care for that disease. Proteomic technologies for finding such proteins study the structure, function and interaction of a specific protein or class of proteins (e.g. kinases), so-called 'functional proteomics', or the global expression levels of a proteome: 'expression proteomics'. Much effort has been made to develop high-throughput analyses, resulting in protein microarrays for functional and expression proteomics [6]. These assays require, to a certain extent, the identities of proteins of interest for choosing suitable capture molecules. Capture is traditionally antibody based, requiring a specific antibody for each protein of interest. Therefore, new methodologies for the faster production of many distinct antibodies and other types of capture molecule are emerging [7].

Functional proteomics assays and protein microarrays can reveal only changes in targeted proteins. However, a 'blind' search of the proteome for proteins that are present in aberrant abundance, called protein profiling, increases the chance of finding proteins with altered expression levels during disease development, progression or treatment. Additionally, protein profiling makes it possible to combine several of these proteins to a discriminative protein pattern. This profiling is most effective when done using protein expression assays in which a large part of the proteome can be analysed at once using a high throughput of samples.

Until recently, two-dimensional polyacrylamide gel electrophoresis (2DPAGE) has dominated this field [8]. A major disadvantage of this technique, however, is its lack of real high-throughput capability. Other techniques used for the expression analysis of proteins are matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS, SELDI-TOF MS, liquid chromatography combined with (tandem) MS (LC–MS–MS), and more-quantitative techniques such as isotope-coded affinity tags (ICAT™) and isotope tags for relative and absolute quantification (iTRAQ™) [5,9–11]. These MS techniques are particularly important for the low molecular weight fraction of the proteome because, in this part of the proteome, the use of immunological assays such as ELISA is limited owing to difficult antibody production for low molecular weight proteins. The advantages and disadvantages of these methods for large-scale protein profiling are summarized in Table 1. The technologies are complementary in their applicability but they all suffer from a limited dynamic range for protein detection. That is, the abundance of serum proteins spans ten to twelve orders of magnitude (from albumin to the low-abundance interleukins) but these methods can, in most cases, detect proteins over only two to four orders of magnitude [12]. This is why low-abundance proteins pose a challenge, especially without prior fractionation of samples.

We focus on the ProteinChip® array-based SELDI-TOF MS [13]. Technically, it is a variant of MALDI-TOF but the on-chip purification is a great advantage. Twelve eight-spot chips are assembled in 96-well 'bioprocessors', which enhances the expression analyses of many samples simultaneously. The fraction of the proteome bound to the chips can be analysed with MS on the same chip, resulting in a 'pattern' of proteins characterized by mass and charge [mass-to-charge ratio ($m/z$)] (Figure 1). Furthermore, the technique is especially suitable for analysing the low molecular weight proteome. Platforms for direct on-chip sequencing of detected biomarkers by quadrupole time-of-flight MS have also become available, thus increasing the feasibility of the identification of discriminative proteins [14]. Consequently, SELDI-TOF MS has excellent potential for protein profiling.

## SELDI-TOF MS applications in clinical oncology

Many reports have described the application of SELDI-TOF MS in cancer and other diseases since the first use of this technique for early-stage ovarian cancer [15], and some of the discriminating proteins have been identified (Table 2). Although SELDI-TOF MS is equally useful for the analysis of cell lysates from tissue and cell lines, in clinical practice most is expected from its application to easily accessible body fluids such as serum. Therefore, we focus on the protein profiling of serum for several important cancer types.

### *Ovarian cancer*

The early detection and treatment of ovarian cancer are crucial for long-term survival. Cancer antigen (CA)125, which is a cell-surface glycoprotein that is abnormally produced and released by ovarian cancer cells, is a currently used serum marker. Unfortunately, its levels are also elevated in some other cancers and in certain benign conditions; therefore, it lacks sensitivity and specificity for diagnosis and is suitable only for treatment monitoring.

Several reports have been published that mention a high sensitivity and specificity of ovarian cancer biomarkers discovered using SELDI-TOF MS (e.g. Refs [16–19]). Criticism regarding the reproducibility of protein profiles and the identity of the proposed biomarkers (tumour originating or due to epiphenomena) [20–24] has emphasized the importance of validation and reproducibility in multiple sample sets and in different laboratories, in addition to the need for biomarker identification. In a multicentre study with cross-validation among 503 individuals from participating