Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Chromatography B, 803 (2004) 27–40

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Methodological advances in the discovery of protein and peptide disease markers

Rainer Bischoff[a,*], Theo M. Luider[b]

[a] University Center of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands
[b] Laboratory of Neuro-Oncology, Department of Neurology, Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract

The quest for biomarkers has seen a renaissance due to the application of newly developed separation methodologies and advances in biomolecular mass spectrometry. It can be argued that each disease influences the physiology of an organism and that these changes should be measurable. Many diagnostic and therapeutic decisions are supported by measurable biochemical or cellular changes in plasma, serum or urine but it is unquestionable that there is a great lack in better markers for early disease detection and prevention. In this review we cover recent developments in the areas of separation science, sample preparation and mass spectrometry as applied to biomarker discovery. We focus, in particular, on the use of LC-MS and SELDI-TOF-MS as two approaches that have seen an upswing in recent years. While validation of newly discovered biomarkers or biomarker patterns and their introduction into diagnostic practice will be a long process, it is our believe that many future diagnostic tests will be based on markers discovered through novel profiling technologies as those outlined in this article.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Reviews; Biomarker; Proteomics

Contents

1. Introduction 27
2. Methodological approaches 29
   2.1. Two-dimensional gel electrophoresis (2DE) 30
   2.2. Liquid-chromatography mass spectrometry (LC-MS) 30
   2.3. Fourier-transform mass spectrometry (FT-MS) 32
   2.4. Surface enhanced laser desorption ionization-time of flight (SELDI-TOF) mass spectrometry 36
3. Final remarks 37
Acknowledgements 39
References 39

1. Introduction

Most diseases manifest themselves by more or less severe changes in human physiology. This forms the basis for clinical chemistry and its value in helping to diagnose disease correctly and in following therapeutic interventions. Presently, many biochemical and cellular parameters are routinely measured in blood, plasma, serum or urine in any major hospital and the results of these measurements support decision making by clinicians (Table 1). One of the major challenges in using body fluid analyses for diagnostic or therapeutic purposes is that there is a large natural variation in the concentration of proteins, peptides or metabolites between individuals ranging from 1 to 40%. Coefficients of variation (CV) within a given assay are considerably

* Corresponding author. Tel: +31-50-3633338; fax: +31-50-3637582.
E-mail address: r.bischoff@farm.rug.nl (R. Bischoff).

1570-0232/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2003.09.004

Table 1
Biochemical and cellular parameters measured by clinical chemistry in body fluids

Assay	Concentration	Mean CV inter-assay (%)	Mean CV intra-individual (%)
Sodium	140 mmol/l	1.0	0.6
Potassium	4.0 mmol/l	1.5	4.8
Calcium	2.50 mmol/l	1.5	1.8
Urea	5.0 mmol/l	2.0	12.6
Creatinine	100 umol/l	2.0	4.4
Uric acid	0.50 mmol/l	2.0	8.4
Iron	30 umol/l	3.0	19.8
Ferritin	15–280 ug/l	5	10
Glucose	10.0 mmol/l	2.5	4.4
Total protein	70 g/l	2.0	2.8
Albumin	40 g/l	2.0	2.8
TSH	0.3–3 mU/l	5	10
ASAT (GOT)	50 U/l	2.5	14.4
ALAT (GPT)	50 U/l	2.5	27.2
LDH	500 U/l	2.5	7.8
Alkaline phosphatase	150 U/l	2.5	6.8
Gamma-GT	50 U/l	2.5	12.0
CPK	150 U/l	3.0	42.0
CPK-MB	0–12 U/l	10	30
Amylase	200 U/l	3.0	9.0
Bilirubin	20 umol/l	3.5	22.6
Triglycerides	2.00 mmol/l	2.5	23.0
Cholesterol	5.0 mmol/l	2.5	5.2
HDL-cholesterol	1.0 mmol/l	4.0	10.0
IgG	8.0–15 g/l	5	3
IgA	1.1–3.7 g/l	5	3
IgM	0.5–2.0 g/l	5	3
BSE	2–15 mm/u	10	10
Hemoglobin	8.7–11.2 mmol/l	1	3
MCV	88–100 fl	2	1
MCH	1.70–2.20 fmol	2	1
Erythrocytes	3.50–5.90 × 10E12/l	2	2
Leucocytes	4.0–10.0 × 10E9/l	3	10
Thrombocytes	150–400 × 10E6/l	3	10
APTT	30–40 s	5	5
PT	10.0–13.0 s	3	5
Fibrinogen	2.0–4.0 g/l	7	15
Homocysteine	8.4 umol/l	4.3	7.0

smaller by comparison. Intra-individual differences result from widely varying nutritional habits, the effect of natural cycles (e.g. day/night, menstrual cycle) and genetic differences between individuals (e.g. CYP450 profiles). Nevertheless, long-term experiences show that many biochemical and cellular parameters can be successfully used to assess disease development and therapy.

When attempting the discovery of new and possibly more predictive biomarkers, it is critical to assess the natural concentration range of such markers and the variation within a given group of individuals. Validation of a given biomarker or multiple correlated biomarkers (marker patterns) is a lengthy process, since the biochemical measurements need to be assessed in light of established clinical criteria, which are often obtained only after considerable periods of time and with great efforts. An example is the assessment of the development of pulmonary emphysema, a slowly progressing destruction of lung tissue, which leads to a reduction in lung function after many years of chronic inflammation. Any biomarker that should allow early diagnosis and assist in defining the most effective therapy will have to be correlated to this slowly progressing disease picture. Similar requirements are needed for other major diseases in the Western World such as atherosclerosis or diabetes type II. Consequently, validation of biomarkers, requires the study of large patient populations over considerable time periods. Finally, a threshold value needs to be defined above (or below) on which action is taken. Threshold values are often further adjusted based upon growing experience over the years and their level may depend on ethnic differences or individual factors. Suitable biomarkers should have a high predictive value and produce a low false positive and negative rate. Thus, there is an interplay between the set threshold level and the generation of incorrect diagnoses as outlined in Fig. 1.

Focus in biomarker research is shifting from methods that can analyze one marker at a time to so called profiling methods, which allow the simultaneous measurement of a range of markers. It is believed that such marker patterns will

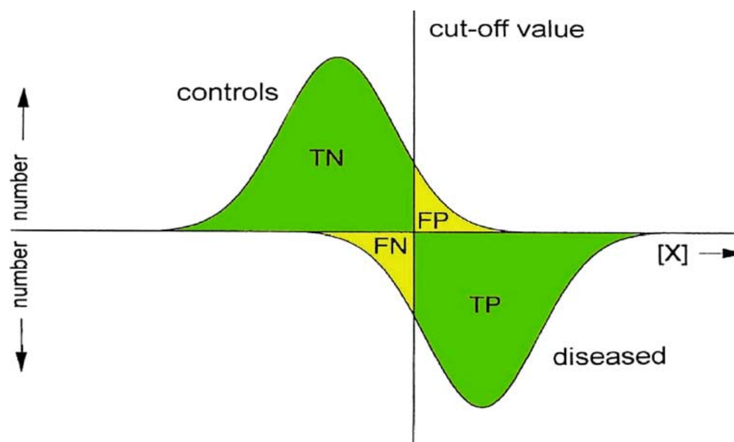

Fig. 1. Schematic illustration of the interplay between setting a cut-off value (threshold) and the possibility of making false negative (FN) or false positive (FP) diagnoses. TN: total negatives and TP: total positives.

allow a statistically more stringent differentiation and a better classification of patient groups. This should improve early detection of disease and also reduce the number of false positive or negative results. Better classification of patients will also diminish the number of subjects that need to be enrolled in clinical trials in order to reach statistically significant levels. Most importantly, it is hoped that a more focused diagnosis will lead to better therapies as well as to earlier intervention.

Successful drug discovery and development relies on measurable criteria to assess efficacy and safety. Such criteria are presently often based on rather complex readouts that are not directly related to the mechanism of action of the respective drug candidate. For example, assessing the effectiveness of a protease inhibitor for chronic obstructive pulmonary disease (COPD) may be assessed by following its effect on the forced expiratory volume after one second (FEV_1), a parameter, though relevant it is, which may only show a response after many years of treatment. This is one of the reasons why research-intensive pharmaceutical companies invest in the discovery of better biomarkers to speed up clinical development and to increase the success rate of preclinical research.

Most of the recent biomarker research has focused on various forms of cancer. Cancer refers to a range of diseases characterized by uncontrolled growth of tissue. Cancers can be classified based on their location or origin or based on the underlying molecular mechanisms of growth. Recently, modern biomolecular, analytical techniques have emerged that promise to improve cancer diagnosis, classification, prognosis, and follow-up of therapy and eventually raise possibilities for individual treatment. All molecular analytical methods are based on the assumption that a growing cancer will affect the physiology of the organism to such an extend that measurable changes will result and thus allow the detection of markers. A presumption is that these changes will result in the active or passive secretion of marker proteins or peptides into body fluids. Marker detection and correlation with tumor growth will be easier for advanced tumors but the true value of novel methods lies rather in early tumor diagnosis, where successful therapy is still possible. Thus biomarker discovery in cancer has the following principle goals:

- Early diagnosis for risk assessment to aid in tumor prevention.
- Early diagnosis that may allow timely therapeutic intervention and cure.
- Reliability; to avoid false positive and negative results in a most acceptable way.
- Follow-up on anti-cancer therapy, specifically the question of tumor relapse.
- Better classification of tumors based on molecular markers allowing a more targeted therapy.

In the following review we will focus on the use of modern analytical techniques, specifically those involving chromatography and mass spectrometry, for the analysis of body fluids in search for novel biomarkers. We will not discuss and review the numerous possibilities of genetic analysis for risk assessment or the use of transcriptomics. The rapidly developing area of bioinformatics can also not be accommodated in this review. For these topics, the reader is referred to the following overviews and publications [1–4].

2. Methodological approaches

Most diagnostic tests are based on blood or urine analyses. In exceptional cases (e.g. breast cancer, medulloblastoma) the analysis of other, less accessible, body fluids is considered such as cerebrospinal fluid (CSF). Blood derivatives such as plasma or serum are often used as starting materials for analysis. In the case of biomarker discovery, it is critical that well-controlled extensive sample collections from patients and controls are available, which is often the case at major medical centers. However, most of these samples were not taken with a proteomics perspective in mind but rather with a view of analyzing established biochemical parameters. Therefore, the first step in setting up a biomarker discovery program is to critically evaluate the quality of the available samples and to get a complete record of their history including the conditions of sampling, transportation, pre-treatment (e.g. coagulation) and storage. This will determine the criteria to include and exclude samples based on technical as well as clinical parameters.

Any methodological approach to biomarker discovery starting with body fluids can be divided into the following “unit operations”:

- Sample banking (e.g. number and size of aliquots, storage conditions).
- Sample pretreatment (e.g. clarification, removal of high-abundance proteins, prefractionation).
- Separation of proteins and peptides (e.g. HPLC, electrophoresis).
- Identification of the separated proteins or peptides (e.g. mass spectrometry).
- Comparative data analysis (e.g. based on quantitative read-outs and pattern comparisons).
- Database development and standardization.

In proteomics of body fluids there are two principally different philosophies dependent on whether the proteins are separated as such followed by proteolytic digestion or whether the whole mixture of proteins in a fluid is digested followed by separation and identification. An example of the “digestion” approach has been used for biomarker profiling in urine [86]. Separating proteins prior to digestion requires methods that work well for complete proteins, such as 2-dimensional gel electrophoresis (2DE), albeit that even this method is not able to display the complete proteome present in a biological sample such as a body fluid and is biased to the high abundant proteins [5]. Proteins

larger than 200 and smaller than 10 kDa are excluded in most 2DE gels. Limited solubility and extremely acidic and alkaline isoelectric points are further characteristics of proteins that are not observed on 2DE gels. Also the concentration range observed in body fluids (six- to nine-orders of magnitude) cannot be covered by this technique. Different solutions have been developed to optimize the reproducibility of 2DE. For instance differential imaging gel electrophoresis (2DE DIGE) uses the possibility to analyze two samples in one gel by using different fluorescent tags [6,7]. However, 2DE remains a low throughput technique and full automation of 2DE is still not commercially available.

Approaches based on chromatographic separations generally require prior digestion of the proteins in a given sample to render them amenable to HPLC, especially if reversed-phase is the method of choice. Even though sample complexity is vastly increased, there is an increasing number of reports on the comprehensive analysis of microbial proteomes using this so-called shotgun strategy [8–14]. To reduce complexity and to avoid overloading of the analytical system, some approaches deliberately eliminate most of the proteins, by for example, concentrating on the lower molecular weight fraction. This facilitates to focus HPLC analysis specifically on the range of small proteins to peptides, a fraction that is also of interest for biomarker studies [15–20]. Mass spectrometry is generally the method of choice for protein and peptide identification. High resolution, high accuracy mass spectrometers such as the Fourier-transform (FT)–MS technique described later, promise to extend the possibilities of this method for biomarker discovery. Although its combination with efficient separation systems is in its beginnings, there are initial results that hold promise [21–23].

Alternatively, there are methodologies that use highly selective molecular interactions to probe the proteome by, for example, immobilizing a range of antibodies on an array, very much like the well-known DNA or oligonucleotide arrays. Using such a method on a proteome-wide comprehensive scale requires, however, that there should be an antibody against each protein present in the sample and that these antibodies need to be highly specific. These requirements are presently not always met despite major advances in antibody and array technology [24–28]. Combining molecular interactions on an array with mass spectrometric detection is a recent addition to the methodological toolbox especially adapted to the rapid screening of clinical samples for biomarker discovery [29–40]. While highly adapted and flexible to the treatment of biological fluids, this method suffers from a lack of identification capacity. The combination of this methodology with mass spectrometers of higher performance (resolution, accuracy, MS/MS capability) may overcome this limitation.

In the following, we will describe a selected range of methodologies in more detail with special reference to their applications for biomarker discovery.

2.1. Two-dimensional gel electrophoresis (2DE)

In proteomics, 2DE is still the cornerstone separation technique for complex protein mixtures although alternatives based on LC-MS and protein arrays are advancing to complement this method. An advantage of 2DE lies in its separation power and the possibility to appreciate directly post-translational modifications. Applications of 2DE to clinical biomarker studies are rare due to the fact that the methodology is not well suited to analyzing large series of samples, as it is quite work intensive especially considering the image analysis part. While powerful software is available for image comparison and clustering, this step requires significant operator intervention and experience partly due to the limited reproducibility of 2DE gels. Recent developments to differentially label samples and to run them on the same gel partially eliminates these problems but it involves a chemical derivatization step, which is not obvious to perform in mixtures containing possibly thousands of proteins [6]. At the end, proteins that are considered to be of relevance as biomarkers are identified mostly on an extra preparative gel according to standard proteomics protocols based on proteolytic digestion and mass spectrometry.

Despite the fact that the approach is work intensive, a number of groups have performed comparative studies over the years and discovered some putative biomarkers [41–43]. However, many tissue samples need to be analyzed and compared to validate proteins that are differentially regulated and that may be used as biomarkers. These initial findings have to be confirmed by specific immunodetection techniques both on 2DE gels and possibly by histochemistry. Although tedious and slow, it may be argued that 2DE has still a strong place in tissue and cell culture analysis (Fig. 2), since liquid chromatography techniques are not highly adapted to separating complete protein mixtures. However, with the advent of integrated multi-dimensional chromatographic approaches of proteolytically digested protein mixtures followed by on-line mass spectrometry and efficient data handling, this picture is changing rapidly [44,85,86]. Application of 2 DE gels to biofluids such as plasma, serum or urine, the major sources of diagnostic markers in the clinic, is less favorable due to the enormous dynamic range of protein concentrations that needs to be covered (approximately nine-orders of magnitude) and the inherent advantage of liquid-based separations for soluble proteins. Furthermore, LC can easily reach into the low-molecular weight protein and peptide range, an area that is largely inaccessible to standard 2DE gels.

2.2. Liquid-chromatography mass spectrometry (LC-MS)

LC-MS is a versatile combination of a commonly used separation technology and mass spectrometry, a powerful identification tool not only for proteins. LC-MS is primarily performed in the reversed-phase mode, since the mobile phase is directly compatible with the requirements for

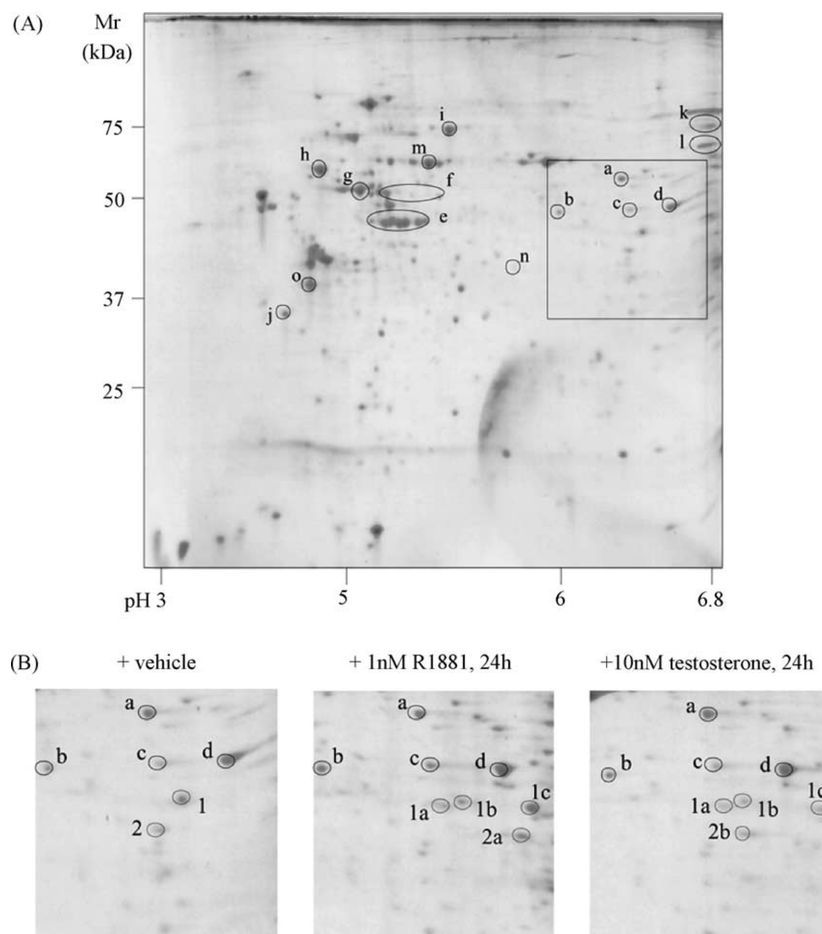

Fig. 2. Androgen-regulated proteins are of particular interest in prostate cancer treatment. By 2DE and MALDI-TOF analysis androgen-related proteins were identified in a cell culture model (MDVD A6) using testosterone, R1881 (a synthetic androgen) and a control vehicle (ethanol) as stimuli, respectively. In panel A the Coomassie brilliant blue stained proteins are shown of cells stimulated with the control vehicle. In the boxed part of the gel differentially displayed post-translationally modified proteins were identified (panel B) as a function of the synthetic androgen and testosterone (1, 1a, 1b, 1c: CarG-binding factor-A in different post-translational modified forms (PTM) and 2, 2a, 2b: PTM of mElfin) (courtesy of Dr. Arzu Umar, from [84]).

ionization at atmospheric pressure. While LC-MS has found widespread use in the analysis of peptides from natural sources or generated by proteolytic digestion of larger proteins, it is not very suitable for analyzing proteins directly. First of all proteins tend to denature under reversed-phase conditions (low pH and high organic solvent concentrations) making their quantitative elution rather difficult and second measuring the molecular mass of a whole protein is not sufficient for its unambiguous identification. For this approach to be successful, it is thus necessary to digest all of the proteins in a sample and to perform the separation at the peptide level. While analyzing biofluids directly without prior digestion is a definitive option in biomarker discovery, prior digestion gives access to the higher molecular weight proteins, however, at the expense of rendering the mixture much more complex. Assuming that a given biofluid contains 1000 proteins and that each protein will give rise to 50 proteolytic fragments, we are talking about separating 50,000 peptides, a goal that can only be approached by multidimensional protein identification technologies [8–14,85,86]. In a large-scale analysis of the yeast proteome by the so-called MudPIT approach it was possible to identify 1484 proteins [13]. Even the low-abundance proteins could be analyzed by this method. A dynamic range of four-orders of magnitude could be reached in the expression level of the proteins present [14]. However, it might be that the order of 1000 proteins in a biofluid is still an underestimate, especially if one includes possible splicing variants and post-translational modifications. This indicates that analysis of such complex mixtures will likely remain a challenge for some time to come no matter which analytical approach is taken. Comprehensive methods, such as 2DE or LC-LC-MS are generally rather time-consuming making them more suitable for initial discovery efforts then for larger clinical validation studies. In the end, it may not be necessary to visualize every low-abundance protein in order to find significant differences that lead to novel markers.

As many reports in the literature have shown, there is still a lot of information in the lower molecular weight fraction of biological samples such as serum or urine. Serum or plasma

may be divided into a high- and a low-molecular weight fraction by ultrafiltration or size-exclusion chromatography. Nevertheless, even after ultrafiltration at a cut-off of 10 kDa there is still a considerable amount of albumin left in urine or in ultrafiltered serum, since the exclusion limits depend on a distribution of pore sizes, which has a certain spread. A combination of size exclusion with adsorptive chromatography is the so-called restricted access chromatography [45]. In this case, only proteins and peptides below a certain size have access to the inner pore surface of the stationary phase and are thus retained while the larger proteins encounter only the hydrophilic, non-adsorptive outer surface and are found in the flow-through. Restricted access materials have found widespread use in the analysis of drug metabolites and other low-molecular weight compounds but have only recently been rediscovered for proteomics applications.

In an example of applying restricted access chromatography to the analysis of hemofiltrate, a biofluid that is comparable to urine, it was shown that the sample preparation step could be performed in-line with the subsequent chromatographic separations [46] (Fig. 3). In conjunction with a two-dimensional chromatographic separation system this allowed to reach a peak capacity of about 5000, a number that is comparable to the number of spots obtained on a highly optimized 2DE gel for proteins. Such high-resolution, multi-dimensional systems promise to provide the methodological means for biomarker discovery in combination with modern mass spectrometers and data analysis software.

Most biofluids contain large amounts of well-known proteins such as albumin and IgGs, which overwhelm the analytical system and may make the detection of the lower abundance proteins and peptides very difficult. It is thus advantageous to remove these proteins prior to digestion and separation. Next to the already described approaches that are based on size, there are alternative ways to reducing the overall protein load by specific adsorption of albumin and IgG to affinity matrices [47–51]. While no affinity matrix is absolutely specific there are degrees of specificity between highly selective immunoaffinity matrices and less selective but more robust affinity supports using synthetic ligands. In an effort to reduce the amount of albumin from human serum, we have evaluated a number of affinity matrices based on antibodies or dye ligands. As shown in Fig. 4 antibody-mediated albumin removal is efficient (Fig. 4A) and selective (Fig. 4B). Dye ligand chromatography, a technique that is extensively used in protein chromatography, is surprisingly effective as well with higher binding capacities and a longer column lifetime, however, at the expense of selectivity.

Application of LC-MS to biomarker discovery is not yet very widespread partly because the method generates large and highly complex data sets that require powerful algorithms and software tools to handle and analyze them. However, these challenges are being tackled so that LC-MS will likely find more use in the future. LC-MS generates information about the components in a sample both in terms of their physico-chemical properties and their molecular mass. Insights given by chromatography such as hydrophobicity or charge of the analytes may provide additional information that can help in sample comparisons. This additional information is, however, presently not systematically used [52].

2.3. Fourier-transform mass spectrometry (FT-MS)

New developments in mass spectrometry are continuously enhancing our possibilities for protein analysis. Fourier-transform-MS (FT-MS) technology provides unsurpassed resolution to derive structural information of large biomolecules in complex protein mixtures [53–55]. For the identification of low-level proteins after digestion into peptides, sensitivities in the low attomole ranges can be achieved using advanced FT-MS technology. In selected cases these sensitivities can be achieved using so called dynamic range enhancement applied to MS (DREAMS) [56] combined with LC columns with bore diameters of 15 μm operated at relatively high pressure (10,000 psi) [57].

FT-MS has a unique specification on resolution (a resolution of 100,000 or more can be achieved) and accuracies can be as good as 1 ppm (both depending on the magnetic field strength). The high resolution, accuracy, the possibility of de-novo sequencing and the high sensitivity give mass spectrometrists the ability to perform specific searches on peptide mixtures obtained from small clinical samples (biopsies, laser dissected parts of frozen tissue sections, and biofluids). Recent developments show the highly intriguing possibility to identify and to quantify hundreds to a few thousands of different proteins in one sample [62].

Latest developments show that because of the unrivaled resolution and accuracy of FT-MS it is possible to cover 40% of the potentially expressed proteins by the human genome with LC/FT-MS and to unravel more than 70% of the potentially expressed proteins in the microorganism, *Deinococcus radiodurans* [56,58]. In addition, information is obtained about the relative amounts of the different peptides observed using different methods to quantify alterations in protein expression by stable-isotope labeling techniques [59–61,87–92].

By FT-MS it was also possible to identify intact proteins in complex assemblies, for example 42 of the 43 core large ribosomal subunit proteins and 58 (of 64 possible) core large subunit protein isoforms were identified in a single analysis [63]. The option of having a MALDI external source present in the newest commercially available FT-MS machines will open possibilities for analyzing arrays and SELDI protein and peptide chips possibly allowing high-throughput, high-accuracy biomarker discovery.

The technology of profiling and direct identification of proteins by using accurate mass and time tags of peptides in complex mixtures of proteins is rather new and until now only successful for micro-organisms. The

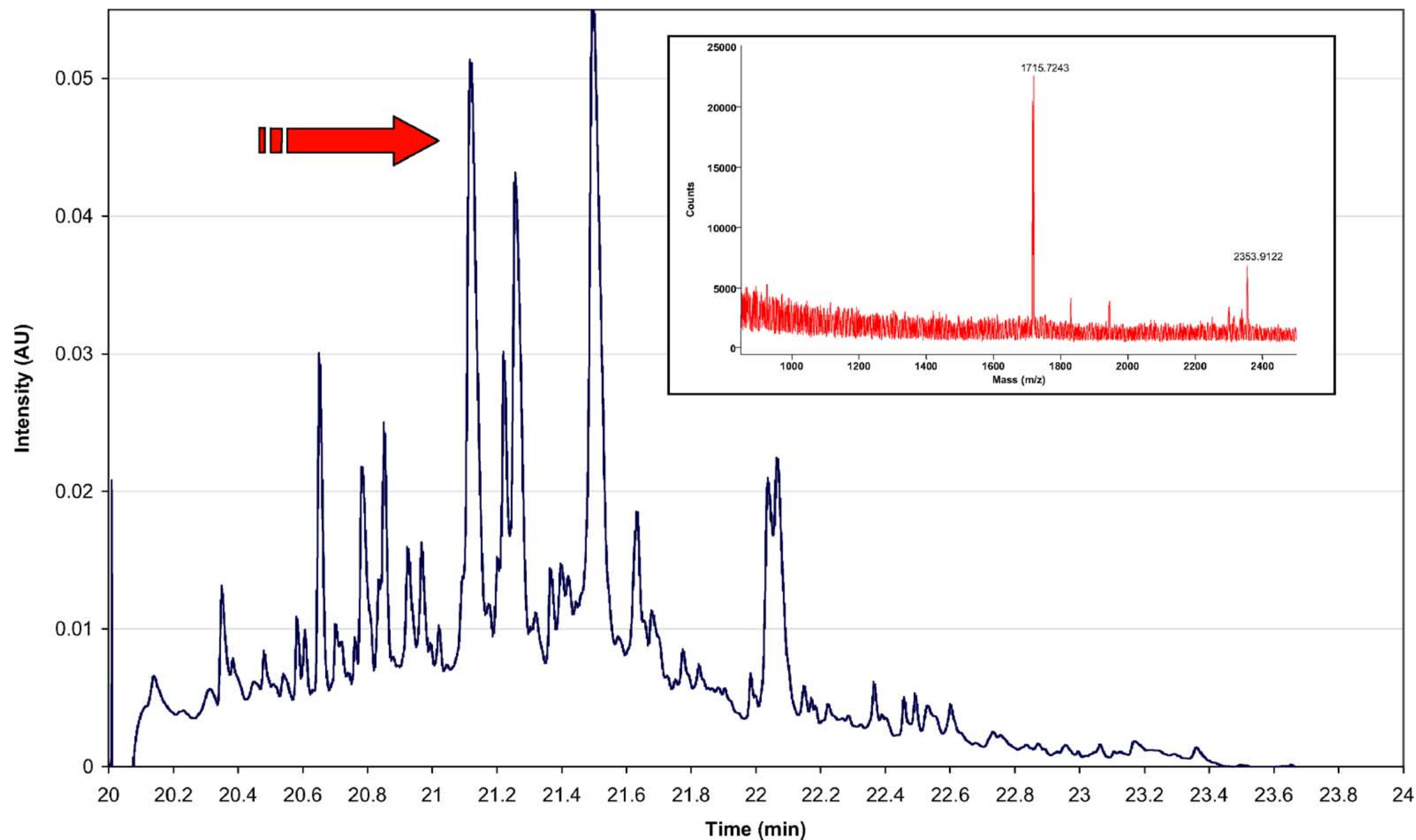

Fig. 3. Multi-dimensional LC-MS: representation of 1 out of 23 chromatograms from the second chromatographic dimension of a modular two-dimensional HPLC system based on cation-exchange and reversed-phase columns in the first and second dimension, respectively. A sample pretreatment step based on restricted access material (RAM) chromatography was included to allow direct analysis of biofluids such as human hemofiltrate. A MALDI-MS spectrum of one of the separated peptides (arrow) is shown in the insert indicating that chromatographic resolution is sufficiently high to detect one major ionizable species in each peak (from [46]).

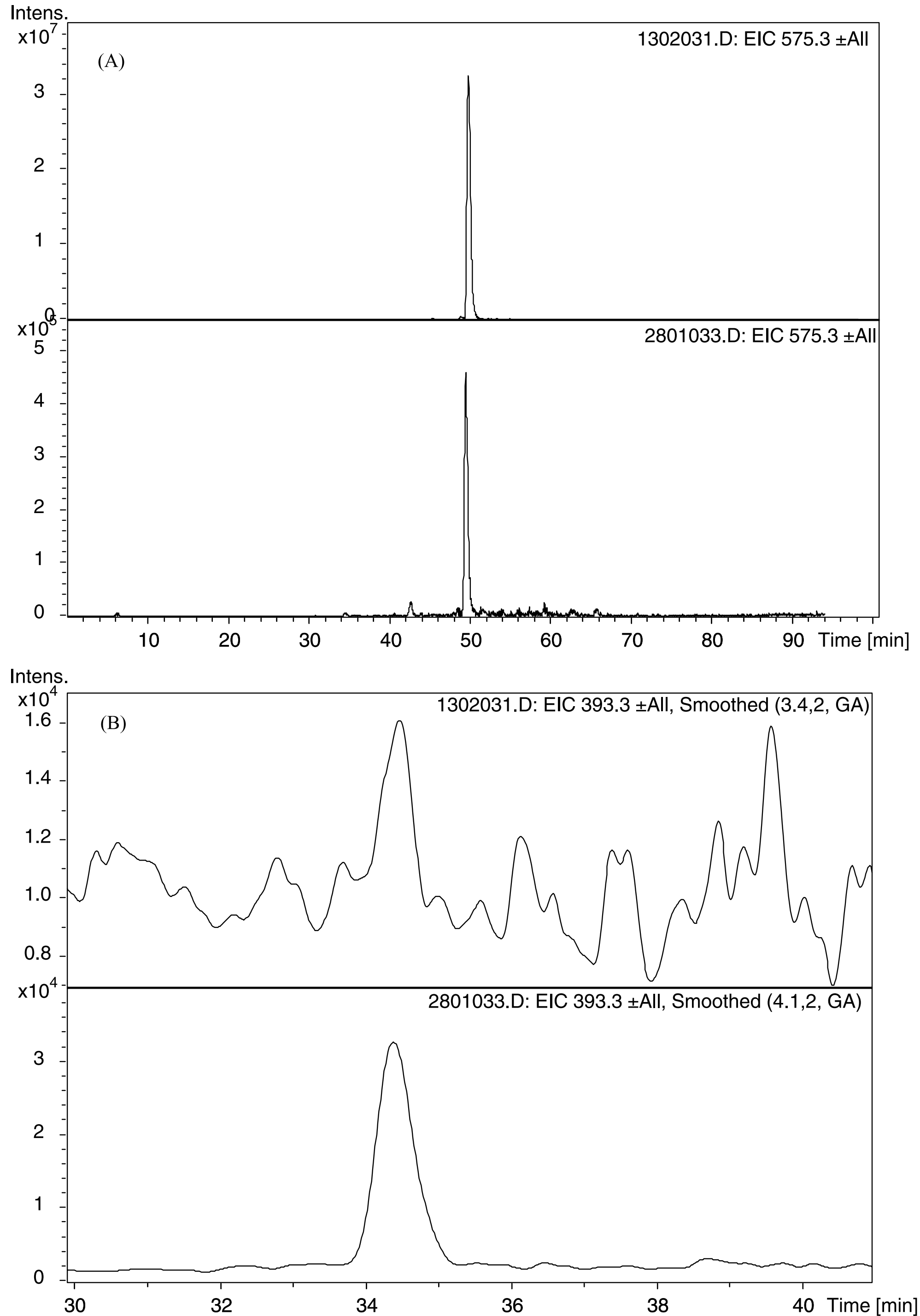

Fig. 4. Efficiency (A) and selectivity (B) of albumin removal from human serum using an anti-albumin immunoaffinity column: (A) extracted ion chromatogram of $m/z = 575.3$ (doubly-charged molecular ion of peptide LVNEVTEFAK; positions 41–50 in human serum albumin) of tryptic digests of human serum (upper trace; peak height 3.2×10^7) or of human serum after depletion with an anti-albumin immunoaffinity column (lower trace; peak height 4.6×10^5). (B) Extracted ion chromatogram of $m/z = 393.3$ (doubly-charged molecular ion of peptide IVDLVK; positions 193–198 in human alpha-1-antitrypsin) of tryptic digests of human serum (upper trace; peak height 16,052) or of human serum after depletion with an anti-albumin immunoaffinity column (lower trace; peak height 32,607). Note the much cleaner detection of this peptide fragment after depletion and the increased overall peak height.

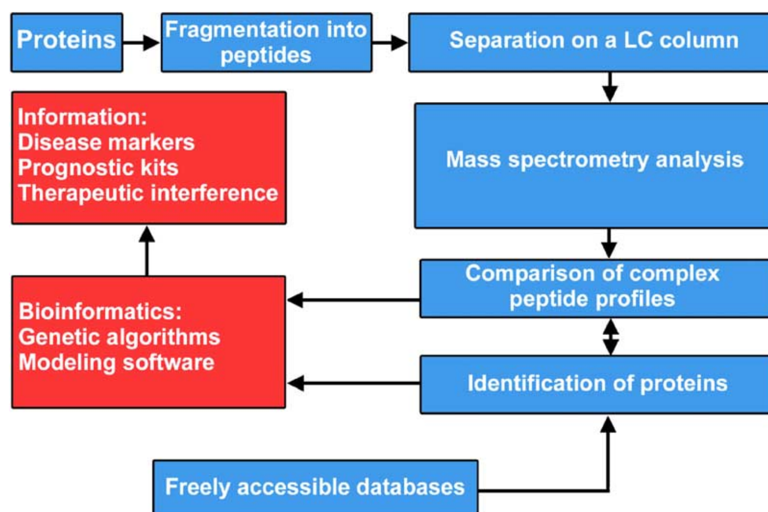

Fig. 5. Schematic view on the use of LC-MS for the discovery of biomarkers in biofluids. The flowchart shows the processing, separation and, analysis of the body fluid samples as well as some of the subsequent mathematical data analysis tools (not reviewed in this article). The eventual comparison and extraction of information is an essential step in the whole process.

possibilities to develop databases with accurate mass tags for eukaryotic organisms is a future possibility with high potential for inter-laboratory biomarker discovery. Because of the higher complexity of these organisms more effort ought to be placed on prefractionation, data handling and the building of large, accessible databases (Fig. 5). In Fig. 6, nanoLC-FT-MS has been used for a comparison of peptide mixtures obtained from cerebrospinal fluid from control and primary brain tumor patients. The high accuracy and resolution of this technology allows to identify peptides and proteins that relate to the primary brain tumor and peptides that are present in the control and the tumor samples just by comparison of the MS spectra. It is to be expected that FT-MS will find further use in biomarker research as commercial instruments become accessible to the wider research community.

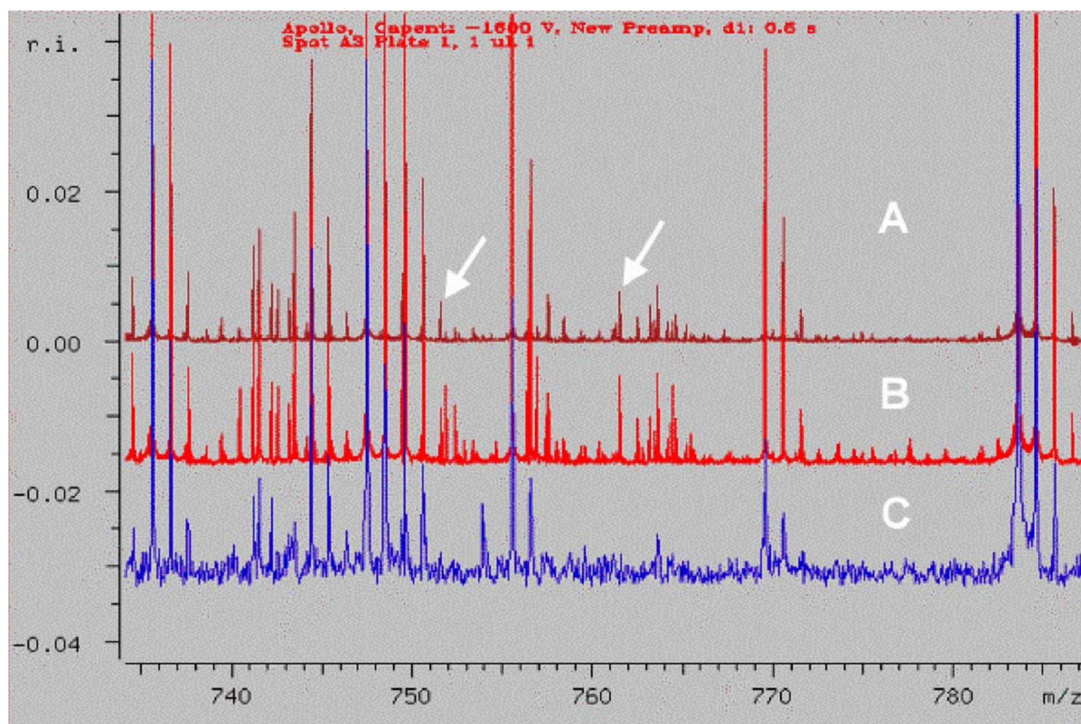

Fig. 6. Examples of relatively small parts of mass spectra obtained by nanoLC FT-MS of a control cerebrospinal fluid sample (blue) and samples obtained from primary brain tumor patients (red). There is a large overlap in all three spectra, however, in the diseased spectra peaks can be appreciated that are absent in the control sample (arrows).

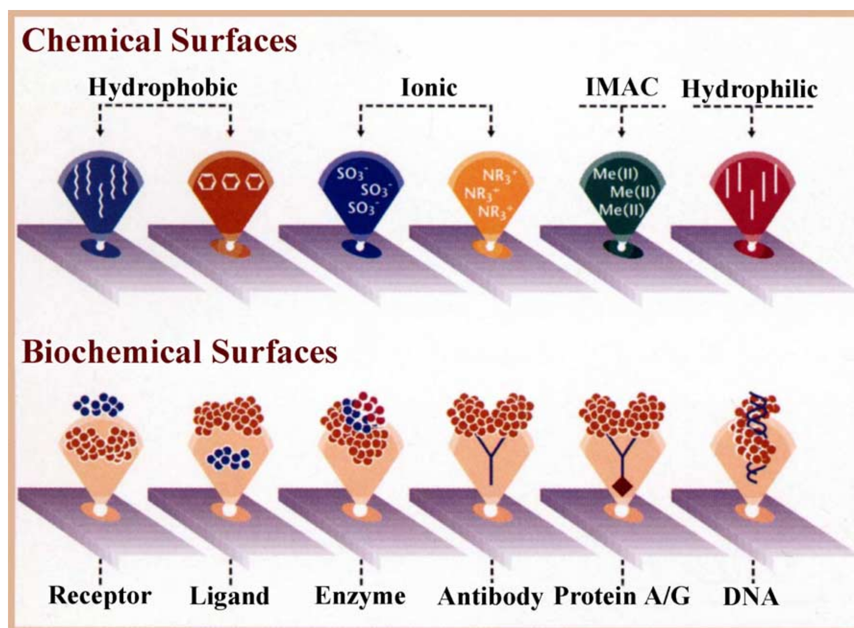

Fig. 7. SELDI protein chip principle: the chips have various surface chemistries so that specific fractions of proteins or peptides bind (Ciphergen Biosystems, Fremont, USA). Molecules that do not bind are removed by washing. After washing a matrix is added and the bound fraction is analysed in a linear TOF mass spectrometer.

2.4. Surface enhanced laser desorption ionization-time of flight (SELDI-TOF) mass spectrometry

Surface-enhanced laser desorption ionization-time of flight mass spectrometry (SELDI-TOF-MS) is an approach that tries to overcome the requirement for purification and separation of proteins prior to mass spectrometry analysis. SELDI-TOF employs a variety of selective chips on which complex bio-materials (e.g. body fluids, cell extracts) can be spotted (Fig. 7). Each of the different chip surfaces will retain a subset of proteins that are subsequently analyzed by a linear TOF mass spectrometer to determine specific and significant pattern changes amongst samples. If specific changes in the expression levels of certain molecules are observed, the nature of these molecules may be determined by mass spectrometers with MS/MS capability. This can be performed in different ways: (a) the proteins of interest are biochemically enriched and purified off line using the knowledge of binding of the protein of interest to the specific chip surfaces or (b) proteins bound to the chip surface are in situ digested, the obtained peptide profiles compared and the peptides of interest directly sequenced by MS/MS after precursor ion selection. By consequence SELDI-TOF allows to compare hundreds to thousands of proteins in a microliter of serum or tissue extract and to process samples rapidly. This has enabled researchers to generate protein profiles and to depict protein-peaks of interest on a much smaller scale. The strength of the MS protein profiles is not the direct protein identification, but the potential to first link the presence or height of multiple individual protein peaks to clinical parameters using dedicated software [64,65]. In this respect, the reliability and reproducibility of the chip and measurement system are of paramount essence. In Fig. 8, an identical serum sample is measured twice on three different chip-strips with the same surface chemistry. One can appreciate the reproducibility on one strip of chips, however, when measuring the same sample on another chip-strip of the same lot number differences are observed (compare graphs A + B and C + D with graphs E + F in Fig. 8). In general a coefficient of variation (CV) of 8–10% ought to be achievable with this kind of technique but in some studies a much higher CV is noted. This indicates that reproducibility of the surface chemistry may still be an unresolved issue with SELDI and that care needs to be taken to check for reproducibility and repeatability [66]. The lack of direct marker identification is a drawback of present day SELDI-TOF technology. To overcome this limitation, we evaluated the possibility to digest proteins bound to the chip. After washing, a specific fraction of peptides remains bound to the chip, which can be sequenced using for instance a hybrid tandem mass spectrometer with an external MALDI source (for review see [67]). Although the spectra are quite similar, small differences can be observed between identical samples processed in the same way indicating that further optimization is needed (Fig. 9). Once peaks are selected and sequenced that distinguish, for example, normal from diseased, one can use these markers in further validation studies. Recently, developed tandem MS technology in conjunction with MALDI is of great interest in this respect [68–70].

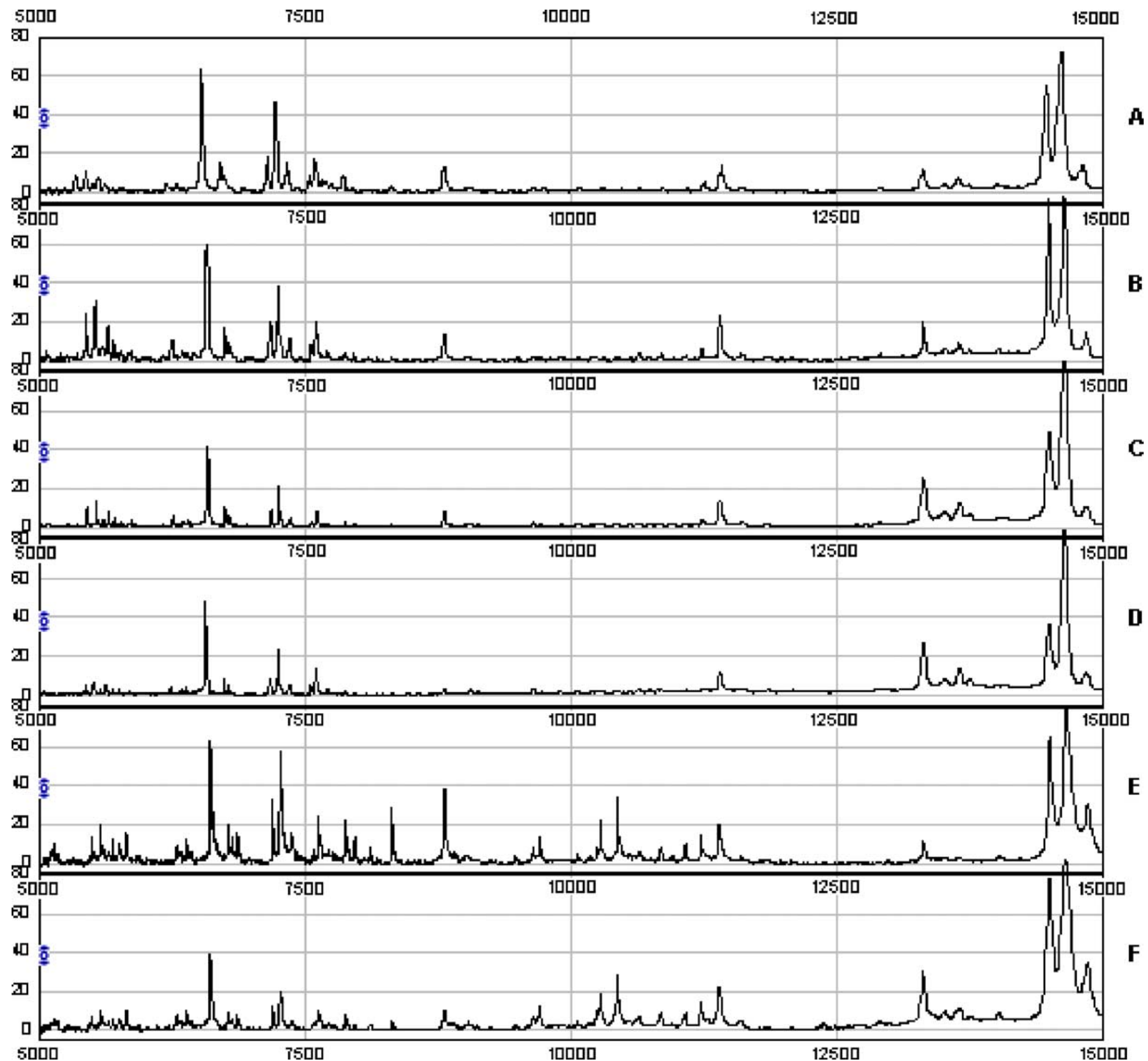

Fig. 8. SELDI protein chip reproducibility: a high interchip-reproducibility is essential for biomarker search by SELDI. In the figure the reproducibility of an identical serum sample is shown on three separate chips of the same batch measured on the same day with identical wash buffers and an identical matrix compound in duplicate. The intrachip-reproducibility on the three chips (A and B) (C and D) and (E and F) is much better than the inter-chip reproducibility (e.g. A and E).

Linking protein patterns to disease has been successfully accomplished for ovarian cancer, breast cancer, prostate cancer and astrocytoma without identifying the respective markers in every case [71–74]. Ovarian cancer was diagnosed with a sensitivity/specificity of 100%/95%, and 83%/97%, respectively [74]. Although comments are raised concerning the claim that population-wide screening can be performed with this technique [75], at present, there is no single diagnostic assay for these cancers with such a discriminatory power. An important feature is the fact that the SELDI-TOF MS analyses were performed on serum, an easily accessible body fluid. A strategy combining surface-mediated protein enrichment with direct mass spectrometric quantification and identification of the putative biomarkers appears to be a promising way for the future even though it is not trivial to correlate observed protein patterns with the purified proteins.

3. Final remarks

Diagnostic and prognostic assays must be cost-effective, reproducible, and uncomplicated. The DNA microarray technology, for example, is rather complex and still expensive and therefore will less likely become a standard assay in diagnostic laboratories. The available patient-derived materials will also enforce restrictions on this type of assays [76]. From all technologies currently available, mass spectrometry on serum to generate complex protein profiles for diagnostic and prognostic evaluation, seems a

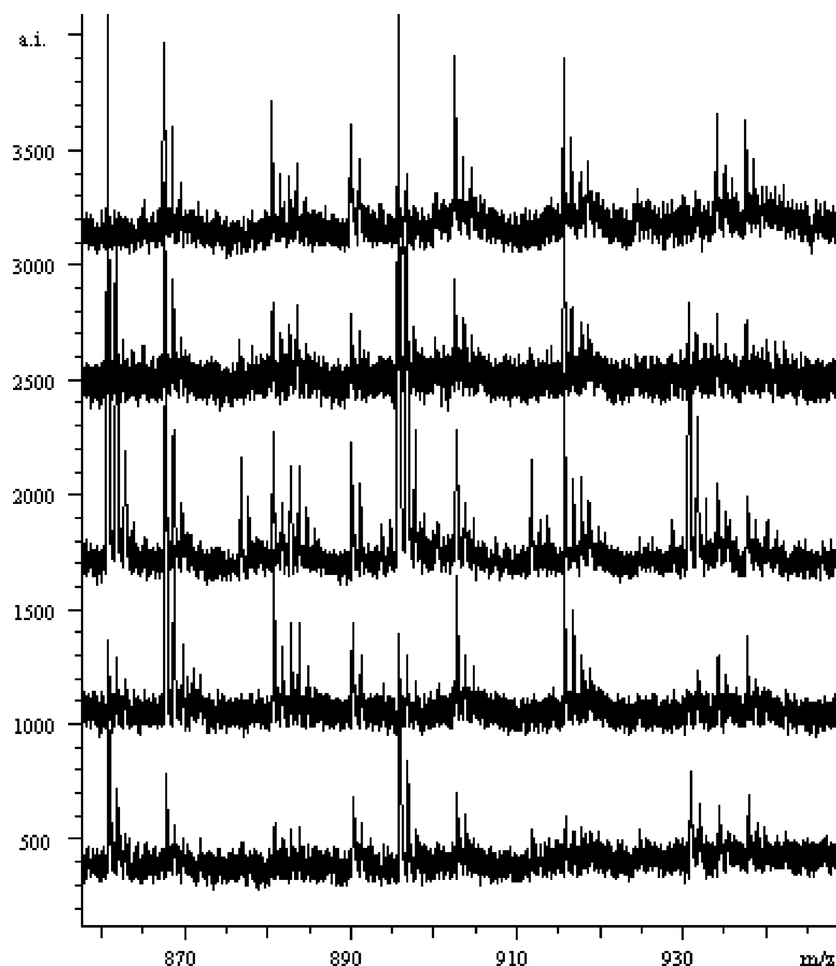

Fig. 9. Generation of peptide maps of proteins adsorbed to a SELDI chip: proteins bound to chips are digested in situ on the chip and the products measured directly in a MALDI-TOF mass spectrometer with the capability of isotopic resolution. Using the MS/MS capability of modern TOF/TOF instruments selected peptides can be sequenced and identified. The graphs shown are five measurements of one sample on three chips of the same batch measured under identical circumstances (one representative analysis is shown).

promising development for clinical implementation. A 1 h, simple and relatively cost-effective assay on the easily accessible serum, could become a valuable tool to reliably diagnose patients, predict disease outcome, and advise on patient-tailored treatment regiments provided that technical hurdles are overcome and standardization is achieved.

New developments in mass spectrometry (improved sensitivity, accuracy, resolution, and relative quantification) and their application in medical research are evolving rapidly. The trend in biomarker discovery is to analyze complex protein and peptide mixtures. The analysis can be limited just to a profile or more advanced to the identification and relative quantification of large series of proteins in one sample. In the nearby future these possibilities will be used not only for research aims but will likely also be applied in medical care. Detection in mass spectrometry can be fast and provide structural information that allows identification, if necessary. This high information content may be combined with the strength of highly specific molecular interactions (e.g. immunoassays) to open further possibilities to develop the protein array technology. While generating molecular diversity to provide a wide range of interaction partners has made great strides in recent years especially based on phage display technologies [77–83,93], protein array technology is still in its infancy. Bottlenecks are the stability of proteins, the numerous different interactions that can occur in complex protein mixtures, and the very high sensitivity needed to analyze these protein interactions, for example in a mass spectrometer.

In this review a number of technical approaches have been outlined that promise to discover novel biomarkers to diagnose disease earlier and to follow up on therapeutic interventions. Biomarker research is high on the agenda of many research groups and especially the pharmaceutical industry. Nevertheless, the timeframe to get to validated biomarkers should not be underestimated and all of those involved will need staying power to succeed. Major advances in analytical chemistry will only lead to benefits for patients if it can be integrated with medical science and bioinformatics.

Acknowledgements

The authors would like to thank the members of the Bischoff and Luider research groups for their contributions. Furthermore, we would like to thank Prof. Frits Muskiet (University Hospital Groningen, The Netherlands) for helpful discussions. Our apologies go to all the many colleagues having made important contributions to this area of research and who's work could not be cited.

References

[1] A. Balmain, J. Gray, B. Ponder, Nat. Genet. 33 (2003) 238–244.
[2] C.M. Michener, A.M. Ardekani, E.F. Petricoin III, L.A. Liotta, E.C. Kohn, Cancer Detect. Prev. 26 (2002) 249–255.
[3] S. Haque, D. Mital, S. Srinivasan, Ann. N. Y. Acad. Sci. 980 (2002) 287–297.
[4] S.M. Ho, K.M. Lau, Curr. Urol. Rep. 3 (2002) 53–60.
[5] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. 97 (2000) 9390–9395.
[6] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[7] W.F. Patton, J. Chromatogr. B 771 (2002) 3.
[8] K. Gevaert, J. Van Damme, M. Goethals, G.R. Thomas, B. Hoorelbeke, H. Demol, L. Martens, M. Puype, A. Staes, J. Vandekerckhove, Mol. Cell. Proteomics 1 (2002) 896.
[9] T.J. Griffin, S.P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood, R. Aebersold, Mol. Cell. Proteomics 1 (2002) 323.
[10] M.J. MacCoss, W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, J.R. Yates III, Proc. Natl. Acad. Sci. 99 (2002) 7900.
[11] W.H. McDonald, J.R. Yates III, Dis. Markers 18 (2002) 99.
[12] W.H. McDonald, R. Ohi, D.T. Miyamoto, T.J. Mitchison, I.I.I. Yates, Int. J. Mass Spectrom. 219 (2002) 245.
[13] M.P. Washburn, D. Wolters, J.R. Yates, Nat. Biotechnol. 19 (2001) 242.
[14] D.A. Wolters, M.P. Washburn, J.R. Yates III, Anal. Chem. 73 (2001) 5683.
[15] M. Raida, P. Schulz Knappe, G. Heine, W.G. Forssmann, J. Am. Soc. Mass Spectrom. 10 (1999) 45.
[16] P. Schulz-Knappe, H.D. Zucht, G. Heine, M. Jurgens, R. Hess, M. Schrader, Comb. Chem. High Throughput. Screen 4 (2001) 207.
[17] P. Schulz Knappe, M. Raida, M. Meyer, E.A. Quellhorst, W.G. Forssmann, Eur. J. Med. Res. 1 (1996) 223.
[18] P. Schulz Knappe, M. Schrader, L. Standker, R. Richter, R. Hess, M. Jurgens, W.G. Forssmann, J. Chromatogr. A 776 (1997) 125.
[19] P. Seiler, L. Standker, S. Mark, W. Hahn, W.G. Forssmann, M. Meyer, J. Chromatogr. A 852 (1999) 273.
[20] E. Machtejevas, H. John, K. Wagner, L. Ständker, G. Marko-Varga, W.-G. Forssmann, R. Bischoff, K.K. Unger, J. Chromatogr. B, this issue.
[21] J.N. Adkins, S.M. Varnum, K.J. Auberry, R.J. Moore, N.H. Angell, R.D. Smith, D.L. Springer, J.G. Pounds, Mol. Cell. Proteomics 1 (2002) 947–955.
[22] J. Bergquist, M. Palmblad, M. Wetterhall, P. Hakansson, K.E. Markides, Mass Spectrom. Rev. 21 (2002) 2.
[23] L. Pasa-Tolic, R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon, R.D. Smith, J. Am. Soc. Mass Spectrom. 13 (2002) 954.
[24] C.A. Borrebaeck, S. Ekstrom, A.C. Hager, J. Nilsson, T. Laurell, G. Marko-Varga, Biotechniques 30 (2001) 11261132.
[25] D.J. Cahill, J. Immunol. Meth. 250 (2001) 81.
[26] B.B. Haab, M.J. Dunham, P.O. Brown, Genome Biol. 2 (2001) RESEARCH0004.
[27] S.P. Lal, R.I. Christopherson, C.G. dos Remedios, Drug Discovery Today 7 (2002) S143.
[28] M.F. Templin, D. Stoll, M. Schrenk, P.C. Traub, C.F. Vohringer, T.O. Joos, Trends Biotechnol. 20 (2002) 160.
[29] B.L. Adam, A. Vlahou, O.J. Semmes, G.L.J. Wright, Proteomics 1 (2001) 1264.
[30] V.E. Bichsel, L.A. Liotta, E.F. Petricoin III, Cancer J. 7 (2001) 69.
[31] E.T. Fung, G.L. Wright Jr., E.A. Dalmasso, Curr. Opin. Mol. Ther. 2 (2000) 643.
[32] M.B. Jones, H. Krutzsch, H. Shu, Y. Zhao, L.A. Liotta, E.C. Kohn, E.F. Petricoin III, Proteomics 2 (2002) 76.
[33] C.A. Nicolette, G.A. Miller, Drug Discovery Today 8 (2003) 31.
[34] C.P. Paweletz, B. Trock, M. Pennanen, T. Tsangaris, C. Magnant, L.A. Liotta, E.F. Petricoin III, Dis. Markers 17 (2001) 301.
[35] C.P. Paweletz, L.A. Liotta, E.F. Petricoin III, Urology 57 (2001) 160.
[36] E.F. Petricoin, K.C. Zoon, E.C. Kohn, J.C. Barrett, L.A. Liotta, Nat. Rev. Drug Discov. 1 (2002) 683.
[37] A.J. Rai, Z. Zhang, J. Rosenzweig, I. Shih, T. Pham, E.T. Fung, L.J. Sokoll, D.W. Chan, Arch. Pathol. Lab Med. 126 (2002) 1518.
[38] P.R. Srinivas, S. Srivastava, S. Hanash, G.L. Wright Jr., Clin. Chem. 47 (2001) 1901.
[39] M. Verma, G.L. Wright Jr., S.M. Hanash, R. Gopal-Srivastava, S. Srivastava, Ann. N. Y. Acad. Sci. 945 (2001) 103.
[40] J.D. Wulfkuhle, K.C. McLean, C.P. Paweletz, D.C. Sgroi, B.J. Trock, P.S. Steeg, E.F. Petricoin III, Proteomics 1 (2001) 1205.
[41] J.E. Celis, P. Celis, M. Ostergaard, B. Basse, J.B. Lauridsen, G. Ratz, H.H. Rasmussen, T.F. Orntoft, B. Hein, H. Wolf, A. Celis, Cancer Res. 59 (1999) 3003.
[42] J.E. Celis, H. Wolf, M. Ostergaard, Electrophoresis 21 (2000) 2115.
[43] B. Seliger, R. Kellner, Proteomics 2 (2002) 1641.
[44] R. Aebersold, M. Mann, Nature 13 (2003) 198.
[45] K.S. Boos, C.H. Grimm, Trends Anal. Chem. 18 (1999) 175.
[46] K. Wagner, T. Miliotis, G. Marko-Varga, R. Bischoff, K.K. Unger, Anal. Chem. 74 (2002) 809.
[47] H.M. Georgiou, G.E. Rice, M.S. Baker, Proteomics 1 (2001) 1503.
[48] A. Kassab, H. Yavuz, M. Odabasi, A. Denizli, J. Chromatogr. B 746 (2000) 123.
[49] K. Nakamura, T. Suzuki, T. Kamichika, M. Hasegawa, Y. Kato, H. Sasaki, K. Inouye, J. Chromatogr. A 972 (2002) 21.
[50] Y.Y. Wang, P.C. Cheng, D.W. Chan, Proteomics 3 (2003) 243.
[51] N.I. Govorukhina, A. Keizer-Gunnink, A.G.J. van der Zee, S. de Jong, H.W.A. de Bruijn, R. Bischoff, J. Chromatogr. A 1009 (2003) 171.
[52] K. Petritis, L.J. Kangas, P.L. Ferguson, G.A. Anderson, L. Pasa-Tolic, M.S. Lipton, K.J. Auberry, E.F. Strittmatter, Y. Shen, R. Zhao, R.D. Smith, Anal. Chem. 75 (2003) 1039.
[53] R.D. Smith, G.A. Anderson, M.S. Lipton, C. Masselon, L. Pasa-Tolic, Y. Shen, H.R. Udseth, OMICS 6 (2002) 61.
[54] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[55] C. Masselon, G.A. Anderson, R. Harkewicz, J.E. Bruce, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 1918.
[56] L. Pasa-Tolic L, M.S. Lipton, C.D. Masselon, G.A. Anderson, Y. Shen, L. Tolic, R.D. Smith, J. Mass Spectrom. 37 (2002) 1185.
[57] L. Pasa-Tolic, R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon, R.D. Smith, J. Am. Soc. Mass Spectrom. 13 (2002) 954.
[58] L. Li, C.D. Masselon, G.A. Anderson, L. Pasa-Tolic, S.W. Lee, Y. Shen, R. Zhao, M.S. Lipton, T.P. Conrads, N. Tolic, R.D. Smith, Anal. Chem. 73 (2001) 3312.
[59] S. Martinovic, T.D. Veenstra, G.A. Anderson, L. Pasa-Tolic, R.D. Smith, J. Mass Spectrom. 37 (2002) 99.
[60] P. Liu, F.E. Regnier, J. Proteome Res. 1 (2002) 443.
[61] M.R. Flory, T.J. Griffin, D. Martin, R. Aebersold, Trends Biotechnol. 20 (2002) S23.

[62] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Veenstra, T..D. Conrads, H.R. Udseth, Proteomics 2 (2002) 513.
[63] S.W. Lee, S.J. Berger, S. Martinovic, L. Pasa-Tolic, G.A. Anderson, Y. Shen, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. U.S.A. 30 (2002) 5942.
[64] G. Ball, S. Mian, F. Holding, R.O. Allibone, J. Lowe, S. Ali, G. Li, S. McCardle, I.O. Ellis, C. Creaser, R.C. Rees, Bioinformatics 18 (2002) 395.
[65] E.F. Petricoin, L.A. Liotta, Trends Biotechnol. 20 (2002) S30.
[66] H.C. Cordingley, S.L. Roberts, P. Tooke, J.R. Armitage, P.W. Lane, W. Wu, S.E. Wildsmith, Biotechniques 34 (2003) 364.
[67] H.J. Issaq, T.D. Veenstra, T.P. Conrads, D. Felschow, Biochem. Biophys. Res. Commun. 292 (2002) 587.
[68] A.L. Yergey, J.R. Coorssen, P.S. Backlund Jr., P.S. Blank, G.A. Humphrey, J. Zimmerberg, J.M. Campbell, M.L. Vestal, J. Am. Soc. Mass Spectrom. 13 (2002) 784.
[69] C.A. Miller, D. Yi, P.D. Perkins, Rapid Commun. Mass Spectrom. 17 (2003) 860.
[70] P. Verhaert, S. Uttenweiler-Joseph, M. de Vries, A. Loboda, W. Ens, K.G. Standing, Proteomics 1 (2001) 118.
[71] A.M. Ardekani, L.A. Liotta, E.F. Petricoin III, Expert Rev. Mol. Diagn. 2 (2002) 12.
[72] D. Carter, J.F. Douglass, C.D. Cornellison, M.W. Retter, J.C. Johnson, A.A. Bennington, T.P. Fleming, S.G. Reed, R.L. Houghton, D.L. Diamond, T.S. Vedvick, Biochemistry 41 (2002) 6714.
[73] A. Wellmann, V. Wollscheid, H. Lu, Z.L. Ma, P. Albers, K. Schutze, V. Rohde, P. Behrens, S. Dreschers, Y. Ko, N. Wernert, Int. J. Mol. Med. 9 (2002) 341.
[74] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L.A. Liotta, Lancet 359 (2002) 572.
[75] Comments on Petricoin et al. (Ref. [74]), Lancet 360 (2002) 169.
[76] J.W. Oosterhuis, J.W. Coebergh, E.B. van Veen, Nat. Rev. Cancer 3 (2003) 73.
[77] P. Amstutz, P. Forrer, C. Zahnd, A. Pluckthun, Curr. Opin. Biotechnol. 12 (2001) 400.
[78] P. Holliger, G. Winter, Curr. Opin. Biotechnol. 4 (1993) 446.
[79] H.R. Hoogenboom, J.D. Marks, A.D. Griffiths, G. Winter, Rev. Fr. Transfus. Hemobiol. 36 (1993) 19.
[80] H.R. Hoogenboom, P. Chames, Immunol. Today 21 (2000) 371.
[81] G. Winter, W.J. Harris, Immunol. Today 14 (1993) 243.
[82] G. Winter, A.D. Griffiths, R.E. Hawkins, H.R. Hoogenboom, Annu. Rev. Immunol. 12 (1994) 433.
[83] G. Winter, FEBS Lett. 430 (1998) 92.
[84] A. Umar, T.M. Luider, C.A. Berrevoets, J.A. Grootegoed, A.O. Brinkmann, Endocrinology 144 (2003) 1147.
[85] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates III, Nat. Biotechnol. 17 (1999) 676.
[86] J.X. Pang, N. Ginnanni, A.R. Donge, S.A. Hefta, G.J. Opiteck, J. Proteome Res. 1 (2002) 161.
[87] M.H. Geng, J.Y. Ji, F.E. Regnier, J. Chromatogr. A 870 (2000) 295.
[88] T.J. Griffin, S.P. Gygi, B. Rist, R. Aebersold, A. Loboda, A. Jilkine, W. Ens, K.G. Standing, Anal. Chem. 73 (2001) 978.
[89] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[90] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.
[91] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6591.
[92] L. Riggs, C. Sioma, F.E. Regnier, J. Chromatogr. A 924 (2001) 359.
[93] E. Soderlind, L. Strandberg, P. Jirholt, N. Kobayashi, V. Alexeiva, A.M. Aberg, A. Nilsson, B. Jansson, M. Ohlin, C. Wingren, L. Danielsson, R. Carlsson, C.A.K. Borrebaeck, Nat. Biotechnol. 18 (2000) 852.