Physiol. Res. 59: 471-497, 2010

REVIEW

Application of Proteomics in Biomarker Discovery: a Primer for the Clinician

V. TAMBOR[1*], A. FUČÍKOVÁ[1*], J. LENČO[1], M. KACEROVSKÝ[2,3], V. ŘEHÁČEK[2,4], J. STULÍK[1], R. PUDIL[2,5]

**These authors contributed equally to this work.*

[1]Institute of Molecular Pathology, Faculty of Military Health Sciences, University of Defence, Hradec Králové, Czech Republic, [2]Charles University in Prague, Faculty of Medicine in Hradec Králové, Czech Republic, [3]Department of Obstetrics and Gynecology, University Hospital Hradec Králové, Czech Republic, [4]Transfusion Department, University Hospital Hradec Králové, Czech Republic, [5]Department of Internal Medicine, University Hospital Hradec Králové, Czech Republic

Received February 3, 2009
Accepted October 5, 2009
On-line November 20, 2009

Summary

Ever since proteomics was proven to be capable of characterizing a large number of differences in both protein quality and quantity, it has been applied in various areas of biomedicine, ranging from the deciphering molecular pathogenesis of diseases to the characterization of novel drug targets and the discovery of potential diagnostic biomarkers. Indeed, the biomarker discovery in human plasma is clearly one of the areas with enormous potential. However, without proper planning and implementation of specific techniques, the efforts and expectations may very easily be hampered. Numerous earlier projects aimed at clinical proteomics, characterized by exaggerated enthusiasm, often underestimated some principal obstacles of plasma biomarker discovery. Consequently, ambiguous and insignificant results soon led to a more critical view in this field. In this article, we critically review the current state of proteomic approaches for biomarker discovery and validation, in order to provide basic information and guidelines for both clinicians and researchers. These need to be closely considered prior to initiation of a project aimed at plasma biomarker discovery. We also present a short overview of recent applications of clinical proteomics in biomarker discovery.

Key words

Proteomics • Biomarker • Plasma • Mass spectrometry

Corresponding author

J. Lenčo, Institute of Molecular Pathology, Faculty of Military Health Sciences, University of Defence, Třebešská 1575, 500 01 Hradec Králové, Czech Republic. Fax: +420 435 513 018. E-mail: lenco@pmfhk.cz

Introduction

During the past decade, several groundbreaking discoveries in life science were made. The completion of sequencing the human genome certainly belongs to the key tasks successfully completed, representing a true milestone in biomedicine (Collins *et al.* 2004). Indeed, this has provided an important knowledge base, thus enabling rapid development in life science-oriented research, in such areas as prenatal and postnatal diagnostics, gene therapy, discovery of new drug targets, and development of personalized therapies (Workman 2003, Lau and Leung 2005, Young *et al.* 2006, Rosa *et al.* 2008). The accomplishment of the complete genome also brings along a new, even more challenging task for scientists: the characterization of the human proteome.

The term "proteome" was used first in 1994 and describes a set of all proteins expressed by a given genome (Wasinger *et al.* 1995). A more accurate definition,

© 2010 Institute of Physiology v.v.i., Academy of Sciences of the Czech Republic, Prague, Czech Republic

emphasizing its dynamic nature, further specifies the proteome as a set of proteins in a given time and space, as its composition may vary from tissue to tissue or even from cell to cell. Furthermore, the structure of a proteome is dependent on a wide range of internal and external factors such as environment, age, sex, diseases, etc., which is in sharp contrast with the nature of the genome.

A protein, the basic unit of a proteome, is a molecule composed of single amino acids, further forming secondary, tertiary, and quaternary three-dimensional structures. Although the amino acid sequence is defined by the appropriate gene, the genetic information itself cannot provide the complete information about a protein. In contrast to the stable, rigid, single- dimensional genomic information based on a combination of four nucleotides, the information encoded in proteins is not exclusively limited to the amino acid sequence. Specific properties of proteins like various conformation states, posttranslational modifications, and alternative splicing demonstrate the multidimensionality, high variability, and dynamic nature of the proteomic information. This explains the high number of unique protein molecules, far exceeding the number of respective genes, particularly in eukaryotes.

Proteomics, the main tool for proteome research, is a relatively new and extremely dynamically evolving branch of science, focused on the evaluation of gene expression at proteome level. Due to the specific properties of proteins mentioned above, current proteomics deals with different issues, such as protein identification, quantification, characterization of posttranslational modification, structure and function elucidation and description of possible interactions. The rapid development of proteomics was made possible by progress in analytical instrumentation, especially in mass spectrometry (MS) with the introduction of new, cutting-edge types of mass spectrometers and improvements of soft ionization techniques. No less important are the advances in technologies and methodologies dealing with protein or peptide separation and sample complexity reduction, mainly in liquid chromatography and electrophoretic techniques. Bioinformatics is the third important foundation for advances in proteomics, as the ability to collect, store, process and visualize vast amount of data is crucial in extensive proteomics studies.

Although genomic research dominated the area of biomedical research in the past decades, proteomics is increasingly gaining ground in leading scientific workgroups and in clinical research labs. One of the reasons driving this platform change is the fact that a protein pattern of a biological sample is much more accurately up to reflecting the current physiological state of an organism than is the genome, and thus holds great promise in biomedicine.

Biomarkers

Timely recognition of an ongoing pathological process is a crucial factor that influences a patient's chances for successful treatment (Etzioni *et al.* 2003, Zhang *et al.* 2007b). To accelerate and facilitate the determination of diagnosis, current medicine strongly relies on the specialized assessment of certain molecules, where the concentration of these molecules in a biological sample more or less correlates with the occurrence of a given disease. Determination of the concentration change of such biomarkers may allow screening of high-risk individuals and detect disease at early, still well curable stages, as well as facilitate the prognosis prediction and monitoring of treatment response. The ultimate goal of implementing these biomarkers in routine clinical tests is the reduction of morbidity and mortality. Unfortunately, even with these tools, it is not always easy to realize the full potential of well-established markers (Andriole *et al.* 2009, Schroder *et al.* 2009).

Requirements of an ideal biomarker

According to the Food and Drug Administration (FDA), a biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic or pathogenic processes or pharmacological responses to a therapeutic intervention. It may be also defined as an *in vivo* derived molecule present at levels deviating significantly from the average in association with specific conditions of health (Atkinson *et al.* 2001, Zhang *et al.* 2007b). From a biochemical point of view a biomarker is often a protein, the presence or quantitative characteristics of which are measured mostly using methods based on monoclonal antibodies. An ideal biomarker should enable unbiased diagnosis determination, particularly in patients without specific symptoms. It should therefore fulfill several criteria, particularly high specificity towards the given disease and high sensitivity. A correlation of the biomarker level and the disease stage is also desirable (Guo *et al.* 2007). Ease of use, standardization, and clarity and readability of the results

for the clinician are all factors that further affect the biomarker performance in the clinical setting. Unfortunately, many of these requirements are not met by most of the potential and even approved and used biomarkers (Anderson 2005). In theory, every disease may be uncovered and characterized by its unique biomarker. To see this biomarker as a single molecule, however, is just one alternative. Rather than as a unique protein, a biomarker should be regarded as a panel of up- and down-regulated proteins or proteins with altered posttranslational modifications, which differ in diseased and normal state (Etzioni *et al.* 2003, Rifai *et al.* 2006).

These facts along with the diagnostic potential of proteins and advances in proteomics technologies recently caused a significant increase of interest in biomarker research. These indicators hold great promise in early detection screening, disease progression monitoring, or in therapy efficiency evaluation, as new, more sensitive and specific markers are yet to be found (Etzioni *et al.* 2003, Veenstra *et al.* 2005, Hu *et al.* 2006, Hanash *et al.* 2008). To illustrate, we present some of recent studies dealing with biomarker discovery, which deserve particular attention because of clinical relevance or biological/methodical approach. These studies are summarized in Table 1.

Biomarker sources

One of the key issues in biomarker research is the accessibility of the source of biological matrix. Among a wide variety of available body fluids, blood is considered the most promising. Other fluids (urine, amniotic fluid, saliva, cerebrospinal fluid, nipple aspirate fluid, synovial fluid, etc.) cannot offer a protein profile as representative as that of blood, and availability of these samples may be very restricted. Blood as a source of biomarkers is easily accessible; its collection is minimally invasive, low risk, and cheap. The processing of crude blood to plasma is a routine task in clinical labs.

Blood

The most important advantage of blood is its contact with virtually all cells of the organism. Due to specific secretion, shedding from the surface, or non-specific leakage, tissue-related proteins are released into the blood stream (Zhang *et al.* 2007a). Therefore, pathologically affected cells with deregulated proteomes may create a specific "barcode" by disease-related proteins released into circulating blood. Besides the proteins originating from affected cells, the barcode is also represented by molecules resulting from organism response to the disease (Bijian *et al.* 2009). Therefore, this barcode includes high-abundance proteins, which can be readily analyzed using conventional techniques. Doubts have emerged, however, on whether these markers would be up to fulfill the criteria required for validation and pass all phases of testing. Except for intact proteins, the barcode also includes protein fragments due to proteases/peptidases deregulation. These are advantageously analyzed using MS profiling (Villanueva *et al.* 2006, Hashiguchi *et al.* 2009). However, the most interesting proteins originate from pathologically affected cells. Unfortunately, owing to the large blood volume, the final concentration of these diagnostically interesting proteins drops to about nanogram per milliliter (ng/ml) levels or even less (Anderson and Anderson 2002, Anderson *et al.* 2004b). To successfully analyze these compounds, sophisticated methods and specific procedures need to be implemented.

Because changes in the plasma proteome are not solely caused by pathological processes, the preanalytical phase is a crucial part of the biomarker discovery workflow. Factors like age, circadian rhythms, stress, medication usage, physical activity, pregnancy etc., may also significantly influence the plasma protein profile. Therefore, all the preanalytical steps – patient preliminary, blood collection, sample transport and storage – need to be strictly standardized and monitored, in order to prevent the occurrence of random and disease-unrelated changes in the plasma proteome. Even minor deviations in the pre-analytical phase may lead to false conclusions of the analysis (Rai *et al.* 2005, Banks 2008, Govorukhina *et al.* 2009). To prevent such deviations, i.e., in blood collection, specialized products like the BD P100 blood collection set (BD Diagnostics, USA) have been developed for proteomic purposes, standardizing the collection procedure. Another crucial aspect, namely, the number of cases and controls enrolled for a study, should also be carefully considered, as an insufficient number of patients may easily lead to false results. For higher credibility, it is advantageous to include patients from multiple clinical centers. In this case, however, strict requirements on standardized sample processing need to be closely monitored, as variations in preanalytical steps may lead even to contradictory results (Fiedler *et al.* 2009).

Table 1. Overview of recent clinical applications of protemics in biomarker discovery projects.

Research area and usefulness	Proteomic platform and validation method	Candidate markers	Ref.
Membranous nephropathy – diagnostic biomarkers	SDS-PAGE of glomeruli protein extract and Western blotting using human sera	Autoantibodies against phospholipase A_2 receptor	(Beck *et al.* 2009)
Hepatocellular carcinoma (HCC) – diagnostic biomarkers	2D-PAGE of HepG2 cells extract and Western blotting using human sera ELISA	11 immunoreactive protein spots were reactive only with HCC sera, among them HSP60 and HSP70	Looi *et al.* (2008)
Chronic inflammatory demyelinating polyneuropathy – diagnostic biomarkers	2D-PAGE (DIGE) analysis of human CSF Nephelometry	Transferrin, α-1 acid glycoprotein 1, apolipoprotein A IV, haptoglobin, transthyretin, retinol binding protein, proapolipoprotein, integrin β 8	Tumani *et al.* (2009)
Lung adenocarcinoma – biomarkers for cancer development and progression	WGA lectin affinity chromatography, 2D-PAGE (DIGE) analysis of human sera Western blot	Adiponectin, ceruloplasmin, cyclin H, proto-oncogene protein kinase Fyn, vanin-2 (GPI-anchored 80-kDa glycoprotein), additional 34 proteins	Hongsachart *et al.* (2009)
Colorectal cancer (CRC) – diagnostic biomarkers	2D-PAGE (DIGE) analysis of human tissue samples Western blot	From 51 tissue protein spots associated with development of CRC, S100A8 and S100A9 were found to be elevated in patients' plasma	Kim *et al.* (2009)
Hepatocellular carcinoma (HCC) – diagnostic biomarkers	SELDI-TOF profiling of human sera ELISA	Peak at *m/z* 13 391 identified as cystatin C, additional 10 peak signatures	Zinkin *et al.* (2008)
Renal cell carcinoma – diagnostic biomarkers	SELDI-TOF profiling of human sera	Peak at *m/z* 8937 identified as eukaryotic initiation factor 2B δ subunit, additional 24 peak signatures	Xu *et al.* (2009)
Melanoma – prognostic biomarkers in early-stage patients	MALDI-TOF profiling of human sera Unspecified immunoassay	Peak at *m/z* 11 680, identified as serum amyloid A, correlating with poor survival	Findeisen *et al.* (2009)
Pancreatic cancer – diagnostic biomarker	MALDI-TOF profiling of human sera ELISA	Three peak signatures at *m/z* 3194, 4055, 5959, and platelet factor 4 represented by peak at *m/z* 7767 and its doubly charged variant at *m/z* 3884	Fiedler *et al.* (2009)
Hepatocellular carcinoma with HCV etiology – diagnostic biomarkers	MALDI-TOF profiling of human sera	Complement C3 peptide, complement C4a peptide and additional four peak signatures	Goldman *et al.* (2007)
Breast cancer – diagnostic and prognostic biomarkers	MALDI-TOF profiling of *N*-glycans released from human plasma glycoproteins	Eight glycan signatures characteristic for breast cancer	Kyselova *et al.* (2008)

Hepatocellular carcinoma (HCC) with HCV etiology – diagnostic biomarkers	MALDI-TOF profiling of *N*-glycans released from human plasma glycoproteins	Three glycan signatures at *m/z* of 2473, 3242 and 4052	Goldman *et al.* (2009)
Chronic allograft dysfunction (CAD) – diagnostic biomarker	LC-MSMS (label free) analysis of human urine peptides LC-MSMS based on Extracted Ion Chromatogram	Uromodulin peptide SGSVIDQSRVLNLGPITR Kininogen peptide DLIATMMPPISPAPIQSDDDW IPDIQI, ions at *m/z* 645.59 and at *m/z* 642.61	Quintana *et al.* (2009)
Rheumatoid arthritis (RA) – diagnostic biomarker	LC-MSMS (label free) analysis of human plasma peptides	Peptides from 25 proteins found differently abundant in patients with RA, peptides derived from thymosin β4 found among the most elevated	Wei *et al.* (2008)
Dilated cardiomyopathy (DC) – diagnostic biomarkers	2D-LC-MSMS (label-free) analysis of mouse tissue Western blot	From 593 mouse tissue proteins associated with development of DC, RTN4 protein found to be elevated in patients' plasma	Gramolini *et al.* (2008)
Breast cancer – diagnostic biomarkers	LC-MSMS (label-free) analysis of mouse tissue MRM, ELISA and Western Blot	Osteopontin and fibulin-2 confirmed as circulating potential markers in mouse model	Whiteaker *et al.* (2007)
Pancreatic cancer – diagnostic biomarkers	2D-LC-MSMS (SILAP) analysis of human sera ELISA	ICAM-1 and BCAM were selected for validation from 121 proteins elevated by factor 1.5 in serum	Yu *et al.* (2009)
Preterm birth (PTB) – screening biomarkers for women at risk	2D-LC-MSMS (SILAP) analysis of human cell lines supernatant MRM for validation in cervicovaginal fluid	From 15 candidates identified in cell line supernatants mixture, desmoplakin isoform 1, stratifin, thrombospondin 1 were confirmed significantly elevated in PTB	Shah *et al.* (2009)
Pancreatic cancer (PC) – early diagnostic biomarkers	LC-MSMS (stable isotope labeling of cystein residues using D0/D3 acrylamide) analysis of mouse plasma ELISA	Five proteins discriminating between patients with PC and healthy individuals up to 13 months prior to development of clinical symptoms	Faca *et al.* (2008)
Endometrial cancer – diagnostic biomarkers	2D-LC-MSMS (iTRAQ) analysis of human endometrial tissue MRM	From nine markers, pyruvate kinase and polymeric immunoglobulin receptor were chosen for subsequent verification and absolute quantification	DeSouza *et al.* (2008, 2009)
Cardiovascular injury biomarkers – previously known markers or marker candidates	MRM, ELISA	CRP, MRP14, MPO, cTnT, cTnI, and NT-proBNP were absolutely quantified in plasma using internal standard.	Keshishian *et al.* (2009)

Although the enormous complexity of blood as a factor reflecting the state of the whole organism may be regarded as an advantage, it may be also seen as a disadvantage from the analytical point of view. Indeed, blood plasma is an extremely rich mixture of proteins and peptides as well as proteins originating from microorganisms. Moreover, proteins may be represented in a number of various forms due to their posttranslational modifications or alternative splicing, which further greatly increases the diversity of the plasma proteome (Anderson and Anderson 2002). Although more than 9,000 plasma proteins have been identified so far, as reported by the HUPO consortium, this was achieved in a collaborative project of 35 laboratories (States *et al.* 2006). Unfortunately, this number of identified proteins is extremely hard to achieve in single-laboratory settings. To illustrate, a more recent work led through very extensive fractionation of serum to the identification of 4,396 proteins in one study (Tucholska *et al.* 2009). The wide concentration range of plasma protein is another limiting factor, as the estimated concentration span exceeds 10 orders of magnitude (Anderson and Anderson 2002). This exceeds the dynamic range of any current analytical instrument or method. The questing for biomarkers thus presents a real challenge for plasma-based proteomics research, as these molecules are hidden among 20 very high-abundance proteins, representing ~ 99 % of total plasma protein (Veenstra *et al.* 2005).

Addressing the problem of high-abundance proteins

In present proteomic research, several methods have been introduced in order to solve some of the pitfalls associated with plasma analysis. One of the key points, often implemented as the first step of proteomic sample workflow, is the removal of ballast high-abundance proteins with no diagnostic potential using immunoaffinity depletion (Tam *et al.* 2004, Echan *et al.* 2005, Huang and Fang 2008). This approach takes advantage of immobilized polyclonal antibodies to remove a portion of high-abundance proteins. These antibodies are designed to bind defined proteins and their isoforms, allowing the removal of up to ~ 95 % of total plasma protein, which results in significant reduction of complexity and dynamic range (Fig. 1). This may lead, in turn, to a higher number of identified proteins, improved sequence coverage, and more accurate protein quantification (Chromy *et al.* 2004, Tam *et al.* 2004, Huang *et al.* 2005b). The depletion step is subsequently included in the validation phase as well, as it enables adequate sample loading (Kim *et al.* 2009). This approach, however, brings along certain disadvantages, as some of the high-abundance proteins, albumin in particular, are known to act as carrier molecules for other proteins, possibly with diagnostic potential. Thus, by removing the carrier proteins, these potentially interesting molecules may be lost as well (Huang *et al.* 2005a, Liu *et al.* 2006).

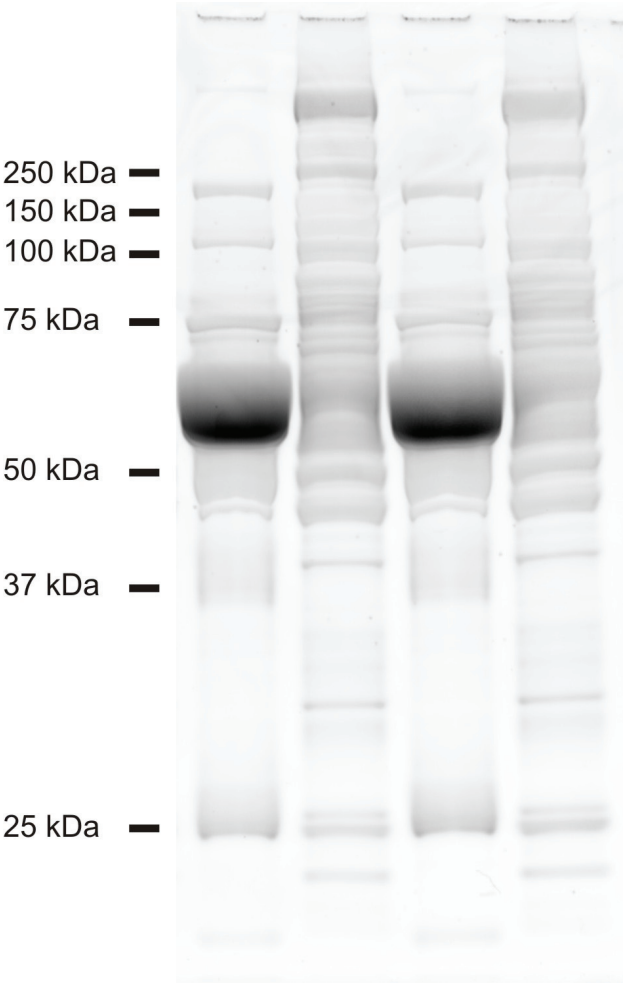

Fig. 1. SDS-PAGE gel of plasma samples processed by immunoaffinity depletion on a MARS Hu-14 column (Agilent). The first and third lane was loaded with the bound fraction, i.e. a fraction containing depleted high abundance proteins. The second and fourth lane present a plasma sample depleted from high abundance proteins.

Peptide libraries present an alternative solution for dynamic range reduction. Instead of removing a portion of high-abundance proteins, the peptide libraries equilibrate concentration of plasma proteins to a similar level. Microscopic beads are covered with a library of hexapeptides prepared using combinatorial synthesis

from common amino acids (Thulasiraman *et al.* 2005, Righetti *et al.* 2006, Righetti and Boschetti 2007, Sennels *et al.* 2007). This results in millions of bead populations, each population carrying a unique peptide sequence. Based on probability, the majority of plasma proteins is supposed to find a binding partner. After the binding capacity of a particular bead population is saturated, the remaining portion of the given protein cannot bind any more and is washed out. The proteins are then eluted from the beads and further analyzed. However, due to the nature of this method, the differences in protein concentration are smoothed among individual samples after saturating the capacity, and only low-abundance proteins that are not up to saturate the beads may be quantified among more samples without employing a method based on stable isotope labeling (Roux-Dalvai *et al.* 2008).

Mining the plasma glycoproteins

The glycosylation of proteins is known to be aberrant in different disease states, especially in cancer (Spiro 2002, Brooks *et al.* 2008). In addition, most of the proteins localized at the surface or secreted by cells are glycosylated. Therefore, disease-related glycoproteins, either actively secreted, or passively shed or leaked from the cells due to cellular damage or death, are likely to occur in the blood stream. Unsurprisingly, numerous clinically used protein markers are glycosylated, such as PSA, CA125, and CEA (Kui Wong *et al.* 2003, Comegys *et al.* 2004, Ludwig and Weinstein 2005, Tajiri *et al.* 2008). Hence, glycoproteomics has been attracting considerable attention in the biomarker discovery field because suitable technologies and methods for glycoproteomic analysis have emerged. With respect to techniques used for this purpose, two approaches can be identified. Lectin affinity chromatography is capable of enriching glycosylated proteins from complex matrices by interaction with various types of lectin without destroying the glycan part, leaving it available for analysis (Mechref *et al.* 2008). During the second alternative approach, covalent capturing of glycoproteins/ glycopeptides, the oxidized glycan moiety is covalently bound to hydrazide solid phase support. The protein/ peptide backbone can be released by enzymatic cleavage from the glycan part and analyzed. The glycan part cannot be recovered from the hydrazide resin and is therefore unavailable for analysis using this approach (Tian *et al.* 2007).

Proximal fluids and tissue

Proximal fluids as a source of biomarkers present a compelling alternative to blood. Although proximal fluids are not as representative as blood, their expedience increases if the nidus of a disease is in close contact with the particular body fluid, i.e., urine may be a prospective source of kidney diseases biomarkers (Quintana *et al.* 2009), or cerebrospinal fluid for central nervous system diseases (Tumani *et al.* 2009). The anticipated biomarker molecules are present in a significantly higher concentration than in body fluids. Moreover, if a disease-specific marker is found in tissue, targeted approaches may be introduced to assess its presence in body fluids as described further in this article (Schiess *et al.* 2009). Unfortunately, the accessibility of tissue specimens or some of the proximal fluids is mostly more complicated compared to collection of blood and may present a level of risk for the patient. Analysis of both tissue and proximal fluids from an animal model of the respective disease may be an option, as these are much more easily obtainable and results from these studies may be then translated to human analogues of the disease (Whiteaker *et al.* 2007, Gramolini *et al.* 2008).

Multistage strategies

Direct analysis of human plasma is just one of many possible ways to seek for new markers. The major obstacles to direct biomarker discovery in plasma – enormous complexity and high concentration range – recently caused several new strategies to emerge (Schiess *et al.* 2009). These are generally divided into multiple parts. First, diseased and control biological samples with anticipated concentrations of potential markers higher than in plasma are compared. These might include model cell lines, affected tissue samples, and proximal fluids (Kulasingam and Diamandis 2008). Apart from the fact that potential markers are present in higher amounts in these sources than in plasma, the overall protein concentration range in cells is lower than in blood, and proteomic analysis of these sources results in higher proteome coverage. Even an animal model may be used, as the diseased and control animals are precisely defined and their genetic backgrounds are alike in all respects (Whiteaker *et al.* 2007, Gramolini *et al.* 2008). Along with the ability to grow a human cancer tissue in the animal host, subsequent analysis allows differentiation of cancer tissue-released proteins from host response proteins (Bijian *et al.* 2009). These are all ways of identifying more specific and sensitive potential

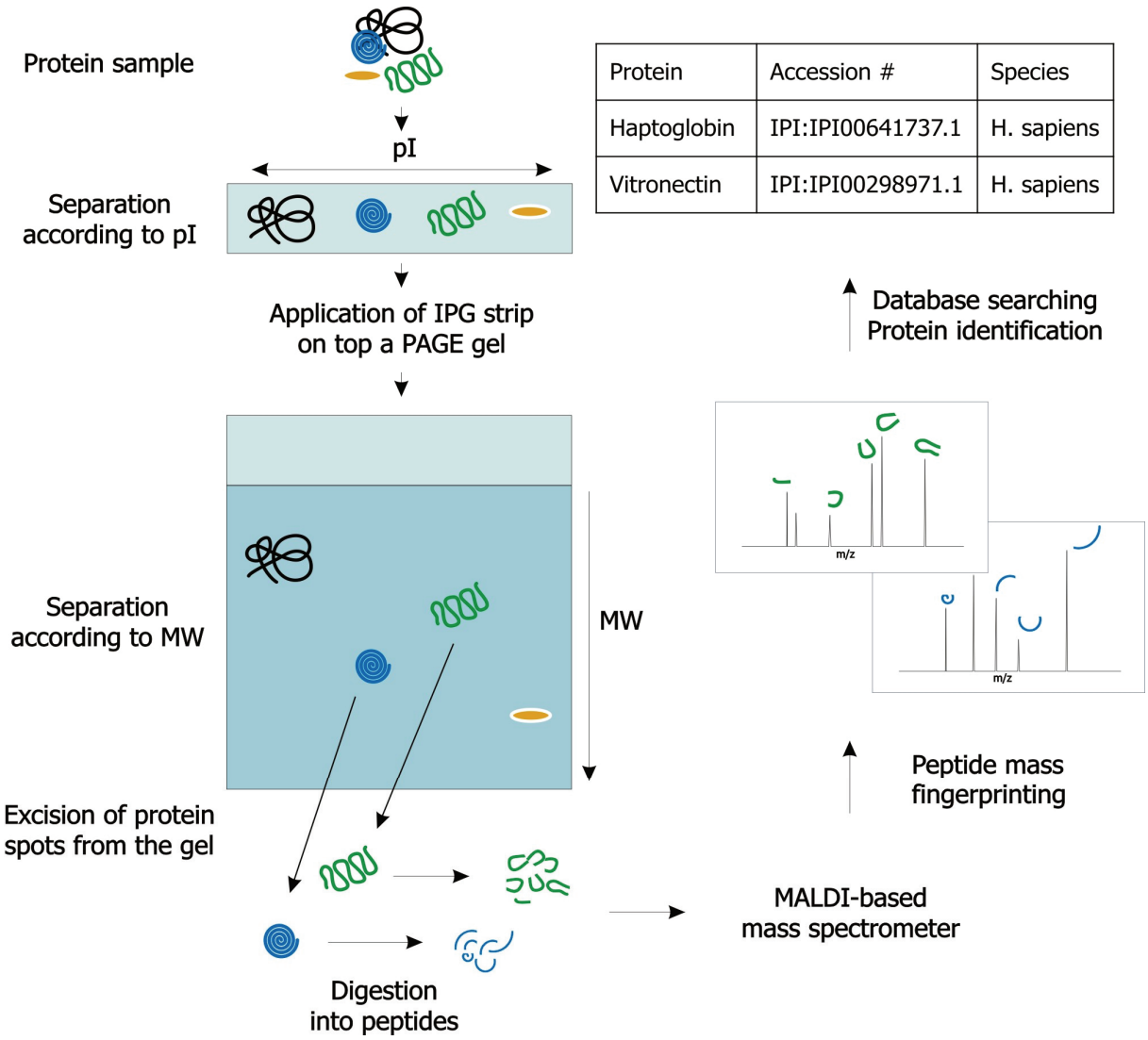

Fig. 2A. 2D-PAGE workflow: A complex protein sample is applied onto an IPG strip and the proteins are separated according to their pI. Then, the strip is placed on top a SDS-PAGE gel and the proteins are separated according to their molecular weight (MW) in second, perpendicular dimension. After gel staining, protein spots of interest may be cut out, digested into peptides and identified mostly by means of PMF approach.

biomarkers. These candidate markers are subsequently targeted in plasma and if their presence is confirmed, then they are simultaneously precisely quantified using targeted mass spectrometry, as described in respective section of this article.

An interesting approach based on identification of glycosylated cell surface proteins was published recently (Wollscheid *et al.* 2009). By this means, it is possible to precisely describe the cell surface proteome. Subsequently, these proteins are targeted using targeted proteomics in human plasma as the cell surface proteins are released into the blood stream upon cellular death or damage. The major drawback of this workflow is the requirement of a suitable representative tissue sample or cellular model of a disease, which is not always available.

Proteomic approaches for plasma analysis

Currently, there are three primary approaches available in biomarker discovery projects (Fig. 2A-C). Each of these methods offers unique advantages but also suffers from specific and often substantial drawbacks. Therefore, one should keep in mind that none of these techniques is ideal and a thorough discussion is crucial prior to selecting the definitive approach. Even though these three methods are fundamentally distinct, a common denominator for all three is the application of mass spectrometry. Therefore, we present a brief description of this key technique.

In principle, mass spectrometry as an analytical technique enables accurate measurements of molecular

weights of individual components in a given sample. A mass spectrometer comprises three major parts: ion source, analyzer, and detector. During a typical MS experiment, sample molecules are ionized and converted into gas phase in the ion source, separated according to their mass to charge ratio in the analyzer, and finally detected by the detector. As for individual segments, the most commonly used ion sources in proteomics are Matrix-Assisted Laser Desorption/Ionization (MALDI) and Electrospray Ionization (ESI). In a MALDI source, analyte molecules are ionized from solid state by a pulsed laser beam, whereas an ESI source ionizes dissolved molecules by spraying them in an extremely fine beam directly into a mass analyzer. For individual types, the commonly used analyzers in proteomics are time-of-flight (TOF), quadrupole (Q), ion trap (IT), Fourier transform ion cyclotron resonance (FT-ICR), and Orbitrap. These might be used either singly or in a tandem configuration. In tandem mass spectrometry (MSMS), multiple steps of mass analysis can be performed with individual analyzers separated in space or in a single analyzer with steps separated in time. In MSMS separated in space, analyzers are physically separated, but are tightly connected in order to maintain vacuum. This configuration is used in the following instruments: Q-TOF, TOF-TOF, Triple Quadrupole, etc. MSMS in time, on the other hand, can be performed with ions trapped in the same place, with individual analysis steps carried out over time. Ion traps or FT-ICRs can be used for this purpose.

Two-dimensional polyacrylamide gel electrophoresis

The very first method employed in comprehensive proteomic experiments was the two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (Fig. 2A). The proteins are separated in a gel matrix based on two independent physicochemical properties of each protein: isoelectric point (pI) and molecular weight (MW) (O'Farrell 1975, Gorg *et al.* 2004, Carrette *et al.* 2006). By a combination of these two features, a high-resolution separation of proteins may be readily achieved.

The protein mixture is separated using isoelectric focusing (IEF) according to the pI of the proteins in the first dimension. The IEF is carried out on commercial gel strips with an immobilized pH gradient (IPG) strips (Bjellqvist *et al.* 1982). The IPG strips containing focused proteins are incubated with sodium dodecyl sulphate (SDS), a detergent that covers the

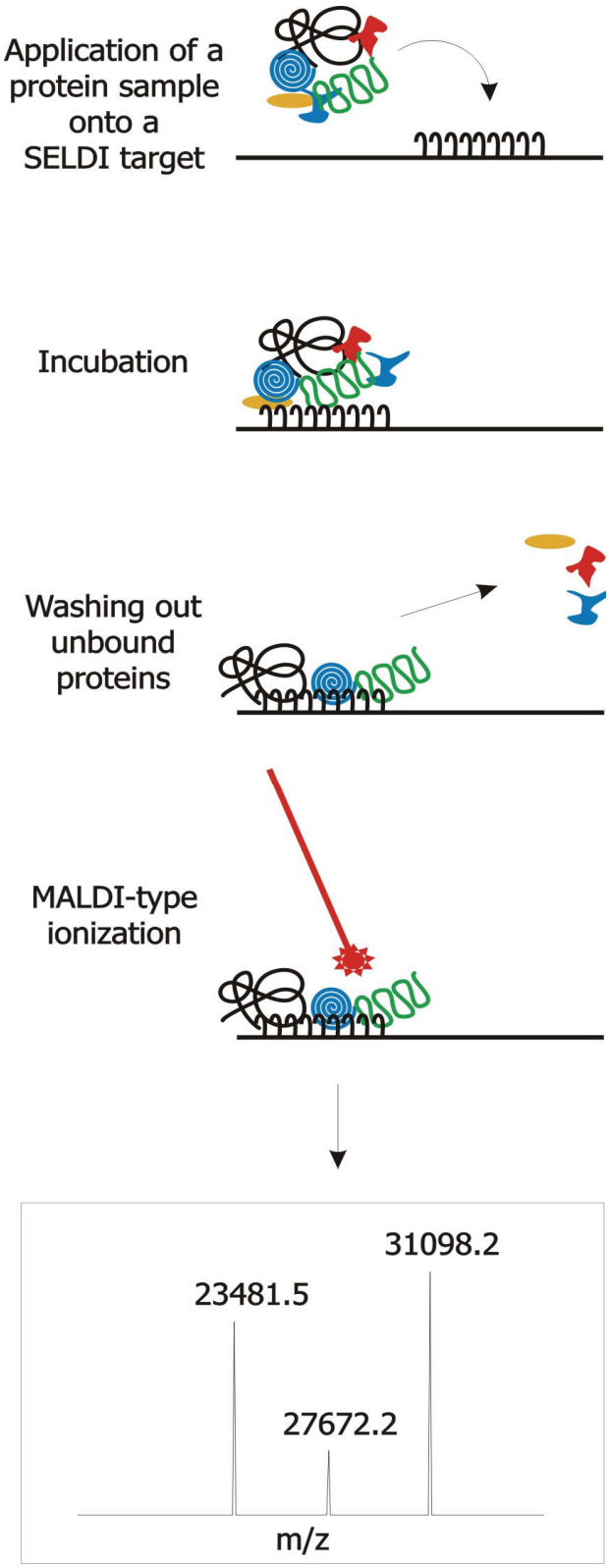

Fig. 2B. Protein profiling workflow: Crude sample is applied onto a SELDI target modified by a specific chromatographic surface. After incubation the unbound fraction is washed away. The SELDI chip is directly analyzed using a SELDI-TOF mass spectrometer. A protein profile is obtained, each protein being represented by a peak with a corresponding m/z value. Note that information on protein identity is missing and cannot be obtained by this type of analysis.

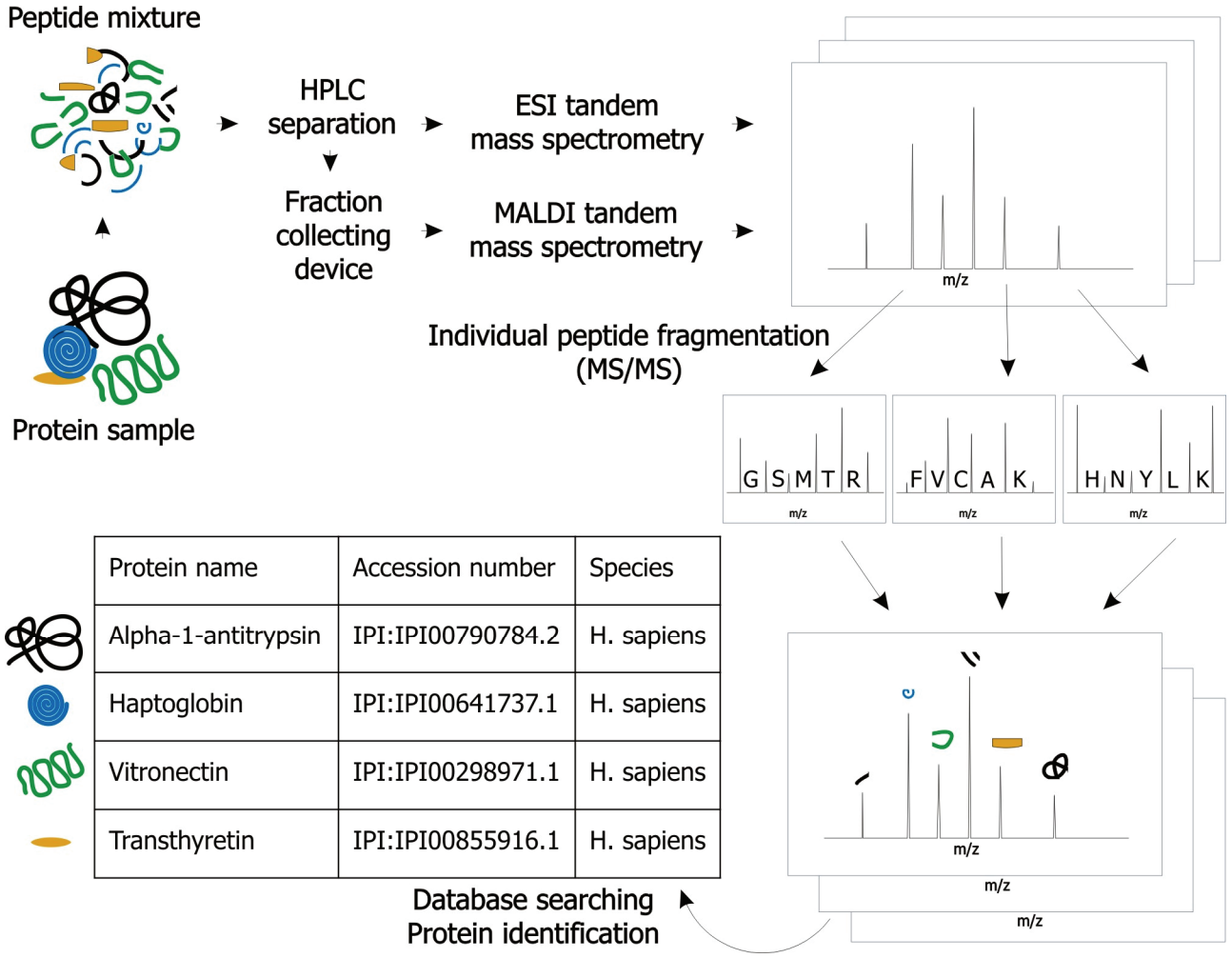

Fig. 2C. Shotgun proteomics workflow: A complex protein sample is digested by a sequence specific protease into peptides. This mixture of peptide may be optionally fractionated and separated. The separated peptides are subjected to MS analysis. First, the MS spectra are acquired and selected peptides from these spectra are fragmented. Resulting MSMS spectra are used for peptide identification. A list of identified peptides is then used in order to identify individual protein components of original sample.

proteins with a negative charge. After incubation, the IPG strips are placed on top of the SDS polyacrylamide slab gels, and the proteins are separated based on their MW in the second, perpendicular dimension (Laemmli 1970). This results in a two-dimensional protein map, where the proteins can be visualized using various approaches.

Conventional staining protocols include Coomassie™ Blue G-250 and R-250 dyes (Neuhoff *et al.* 1988, Candiano *et al.* 2004), or a color reaction based on silver ions reduction of ionic to metallic silver onto the protein surface (Rabilloud *et al.* 1994, Chevallet *et al.* 2008). Increasingly popular fluorescent dyes, e.g. Sypro™ Ruby (Berggren *et al.* 2000) and Deep Purple™ formerly known as Lightning Fast, (Mackintosh *et al.* 2003) offer ameliorated sensitivity and linearity for quantification compared to classic staining agents. The Differential Gel Electrophoresis (DIGE) employs three fluorescent dyes (Cy2, Cy3, and Cy5) for covalent protein labeling prior to 2D-PAGE. Due to their identical physicochemical properties in regard of pI and MW, labeled proteins are run on the same gel simultaneously. However, due to different excitation and emission wavelengths of the dyes, a unique 2D protein map can be acquired for each protein sample loaded on the gel (Unlu *et al.* 1997). When choosing the appropriate staining protocol, factors like sensitivity, dynamic linearity, and compatibility with MS analysis should be taken into consideration (Miller *et al.* 2006, Berth *et al.* 2007) (Table 2). Subsequently, the stained gels are digitalized and evaluated by means of specialized software enabling quantification of proteins via comparison of the intensity of stained spots (Berth *et al.* 2007). This final step is crucial, as any variance in image processing may lead to false results, mostly in quantification (Stessl *et al.* 2009).

Table 2. The most frequent staining methods (based on Miller *et al.* 2006).

Staining method	Principle of detection	Sensitivity	Linearity for quantification	MS compatibility
Coomassie™ Blue G-250 colloidal	Absorption	++	++	+
Silver staining	Absorption	+++	+	–/+
Sypro™ Ruby	Fluorescence	+++	+++	+
CyDyes - DIGE staining	Fluorescence	++++	++++	+

The separated protein spots are identified on a mass spectrometer, mostly using the peptide mass fingerprinting method (PMF) (Shevchenko *et al.* 1996, Henzel *et al.* 2003). A gel piece containing an isolated protein is excised and enzymatically digested by trypsin or any other sequence specific protease, resulting in a mixture of peptides. A MS spectrum is acquired, each peptide being represented by its mass-to-charge ratio (*m/z*) value. The recorded *m/z* values are compared with theoretical values and in case of a match, a protein is assigned to a spectrum with certain probability, according to the Mowse scoring algorithm (Pappin *et al.* 1993). The theoretical *m/z* values are obtained by *in silico* translation of DNA sequences of genes into proteins, from which theoretical proteolytic peptide masses are computed. If a spectrum fails to provide sufficient data for confidential protein identification, a tandem spectrometer may be used, as this type of instrument enables direct acquisition of a peptide sequence (Thiede *et al.* 2005).

Several hundreds to a few thousands of protein spots may be separated on a single 2D-PAGE gel. This approach is one of the most suitable for separating isoforms of identical proteins. Also, the expenditure for the required equipment and chemicals is relatively low. However, the main drawbacks of 2D-PAGE include reproducibility issues, time and labor intensiveness of the process, and imperfect separation of protein in both pI and MW extremities and of hydrophobic proteins. A partial solution to the reproducibility and dynamic range problems may be achieved using the DIGE approach, solving also problems regarding the low dynamic range of conventional staining methods.

Although the 2D-PAGE method has been applied to numerous projects for biomarker discovery, the proteins with altered concentration belong mostly to the group of high-abundance proteins (Tumani *et al.* 2009). However, if specific fraction or enrichment methods are employed during the sample processing workflow, even tissue-derived proteins may be detected using this approach (Hongsachart *et al.* 2009). Nonetheless, gel-based techniques may bring substantial results in a very specific field of biomarkers, namely autoantibodies that act as markers. In autoimmune diseases or in cancers autoantibodies are often found to be targeted against own cellular proteins (Bazhin *et al.* 2009). In this case, the strategy of searching for biomarker is far different from those described above, as the marker itself is an immunoglobulin and the task is to determine against which antigen it is targeted. The strategy is to perform Western blotting of affected tissue proteins by using imunoglobulins from the sera of patients. Although protein-antibody arrays currently dominate this area, conventional gel-based proteomic methods can still bring significant results (Looi *et al.* 2008, Beck *et al.* 2009).

Proteomic profiling (Fig. 2B)

Direct MS analysis of a sample may provide rapid insight into its protein profile. An instrument based on MALDI-TOF in linear configuration is ideal for this purpose, as it enables an acquisition of wide *m/z* range. By this approach, protein profiles of samples may be quickly compared, resulting in a list of differentially concentrated protein peaks (Fig. 3). However, due to the complexity of biological samples, the majority of low-abundance proteins remain undetected. This issue is partially solved by sample prefractionation on a carrier, covered by various chromatographic surfaces. These bind only the desired subset of proteins and the corresponding protein profile is then acquired using a mass spectrometer directly from these carriers. This approach is also known as Surface-Enhanced Laser Desorption/Ionization time-of-flight mass spectrometry (SELDI-TOF MS) (Dattelbaum and Iyer 2006, Poon 2007). Currently, a variety of chemical and biochemical surfaces is at disposal, enabling analysis of a wide range of protein subgroups. Analogous analyses may be also performed

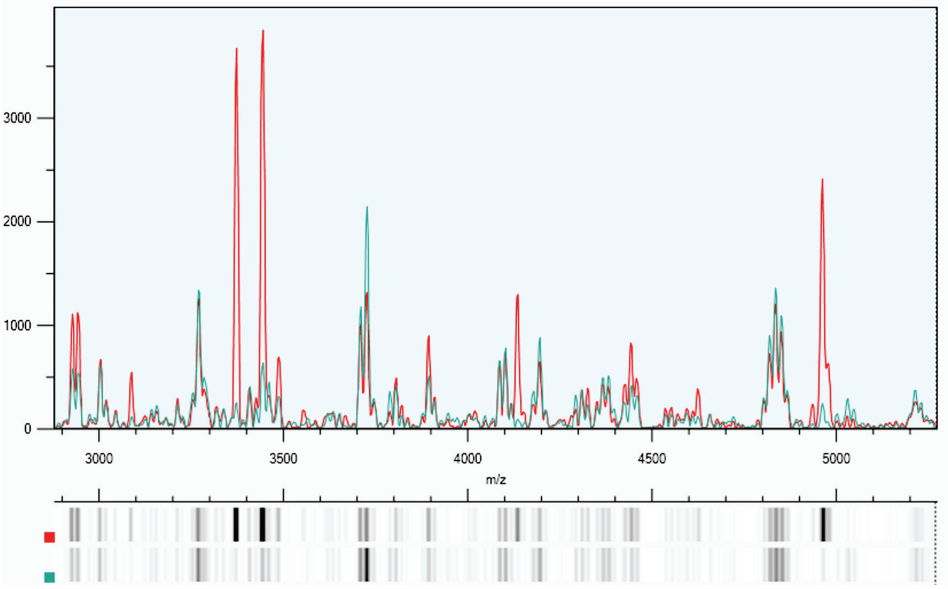

Fig. 3. Representative MS profiling spectrum. Two samples obtained from infection free (shown in green) and infected (shown in red) amniotic fluid were acquired on a MALDI-TOF mass spectrometer in linear mode and compared. Several markedly altered peaks were detected. Except the spectra, an alternative gel-like view is also shown.

on a MALDI-TOF instrument, but the sample prefractionation has to be performed separately, i.e., using magnetic beads modified by various chromatographic surfaces, similar to those on SELDI carriers, or using column devices filled with chromatographic phases. This configuration enhances sensitivity, as the surface of beads is higher compared to those of SELDI targets. Due to the poor analytical performance of SELDI-TOF instruments, researchers experienced in mass spectrometry prefer alternatives based on MALDI-TOF technology for biomarker discovery applications (Villanueva *et al.* 2004, Callesen *et al.* 2009).

Compared to 2D-PAGE, a SELDI-TOF analysis requires a much lower amount of sample, which may be in addition applied directly onto the target, without extensive preparation. Also, this technique is remarkably fast and high-throughput. Nevertheless, the SELDI-TOF approach suffers from some major drawbacks, including low spectra resolution and low accuracy. In addition, concerns about reproducibility discourage MS profiling from becoming a routine proteomic tool prior to addressing standardization of preanalytic and analytic factors (Banks 2008, Bruegel *et al.* 2009, Callesen *et al.* 2009). Furthermore, the absence of means for precise protein identification in SELDI-TOF limits the information about a biomarker candidate protein to just its *m/z* value in most cases. Although publications presenting just these limited data on candidate markers keep emerging, proteins/peptides defined just by *m/z* are worthless for diagnostic applications because their unknown identity hinders further validation by independent orthogonal methods. Limited or no options for this validation step further increase controversy and skepticism currently associated with this approach. Even though the SELDI-TOF technique or profiling based on MALDI-TOF instrument shows some disease-related changes in plasma, these occur mainly among the higher-abundance proteins (Hu *et al.* 2006, Findeisen *et al.* 2009). Due to their low specificity, however, these would unlikely pass the validation for a disease-specific biomarker. On the other hand, as the profiling approach focuses on low *m/z* segment, disease-specific low-molecular weight fragments may be detected in plasma as certain pathologies are characterized by profound deregulation in proteases/peptidases activities (Villanueva *et al.* 2006, Goldman *et al.* 2007, Hashiguchi *et al.* 2009). Another area where the profiling strategy can be advantageously employed involves analysis of glycans from glycoproteins. As already mentioned, the glycosylation pattern of proteins is known to be aberrant in different diseases. One of the methods shown to be able to uncover disease-specific changes in glycosylation is MALDI profiling of *N*-glycan moieties released from plasma/serum glycoproteins. This method has been proved to be well reproducible (Wada *et al.* 2007). To date, sera from various cancer patients have been tested using this approach (Kyselova *et al.* 2008, Goldman *et al.* 2009). Although this approach seems to be very promising, as it has been shown to be able to distinguish individual cancer stages (Kyselova *et al.* 2008), concerns have been raised on how to identify the parent glycoproteins, allowing further validation.

Shotgun proteomics (LC-MSMS) (Fig. 2C)

The combination of liquid chromatography (LC) and MS allows detection of proteomes with greater depth, dynamic range, and enhanced accuracy of quantification than when using one-dimensional profiling techniques that record all ions in a single mass spectrum. The shotgun approach is closely linked to advances and progress in MSMS. A tandem mass spectrometer is an instrument capable of isolating a precursor ion, fragmenting it, and detecting resulting fragments (Domon and Aebersold 2006).

During a typical shotgun experiment, a protein mixture of various complexity is cleaved by a sequence-specific protease first. The most commonly used protease in proteomics is trypsin that cleaves a typical protein into several tens of peptides. Therefore, in case of analyzing a complex protein sample, a huge amount of different tryptic peptides raised from trypsin digestion disallows a direct MS analysis similar to the PMF method. Therefore, the resulting peptide mixture has to be separated, mostly by high performance liquid chromatography (HPLC), prior to analysis on a tandem mass spectrometer. These two systems may be connected either on-line, where the HPLC capillary flows directly into the ESI ionization source, or off-line, using a fraction collecting device. This device collects the peptides eluting from a HPLC system in time-dependent fractions directly onto a MALDI target plate (Bodnar *et al.* 2003). Alternatively, a continuous elution trace may be deposited onto the MALDI plate, which results in increased chromatographic resolution, comparable to that of ESI-based MS instruments (Chen *et al.* 2005).

The mass spectrometer first acquires a MS spectrum of intact peptides, from which candidates are selected for fragmentation. In case a peptide meeting specific requirements on its intensity and charge is detected, this peptide precursor is isolated from the others, fragmented, and the resulting fragments then provide a MSMS spectrum. Information acquired from both MS and MSMS spectra is used to identify of the proteins in the original mixture (Nesvizhskii 2006). At present, several searching tools and algorithms are available. Most of these tools are based on the precursor approach (Mascot, Sequest), which uses the precursor mass value as the main search criterion (Clauser *et al.* 1999) and takes both MS and MSMS spectra equally into consideration. On the other hand, the sequence tag approach is based on partial *de novo* peptide sequencing and uses mainly the acquired MSMS spectra (Mann and Wilm 1994).

The most common peptide separation scheme nowadays is based on HPLC, using a stationary C18 reversed phase (RP) column providing excellent resolution. Along with good separation efficiency, an additional advantage of this method is the use of solvents, which do not inhibit either ESI or MALDI type of ionization (Mitulovic and Mechtler 2006). However, a single dimension RP HPLC is not powerful enough to resolve a mixture of hundreds or thousands of various peptides resulting from an enzymatic digestion of a complex protein mixture like plasma (Gilar *et al.* 2009). Therefore, various fractionation and separation methods are combined to simplify the analyzed mixture as much as possible. One of these combined schemes incorporates a strong cation-exchange chromatography (SCX) HPLC prior to the RP HPLC. This approach was denoted as Multidimensional Protein Identification Technology (MudPIT) (Wolters *et al.* 2001). Alternatively, SCX HPLC in the first dimension may be replaced by RP HPLC in basic conditions (Gilar *et al.* 2009) or even by IEF of peptides providing at least comparable fractionation efficiency (Essader *et al.* 2005). The GeLC-MSMS method combines a SDS protein electrophoresis followed by gel cutting, protein digestion and RP HPLC separation of the resulting peptides (Schirle *et al.* 2003). The HUPO Plasma Proteome Project data unambiguously showed that the shotgun approach using these multidimensional separation methods leads to a much higher number of identified proteins than does the 2D-PAGE approach (Omenn *et al.* 2005). Also, a combination of various fractionation and separation methods leads to partially redundant sets of identified proteins. In general, the more orthogonal methods are combined, the higher the number of identified proteins. On the other hand, along with the fraction count, the analysis lengthens proportionally and the procedure becomes more error prone (Hoffman *et al.* 2007).

Quantitative shotgun proteomics

The main goal of former shotgun proteomic studies was mainly protein identification. However, advances in mass spectrometry and bioinformatics enabled a focus shift towards quantitative and comparative analyses where a comparison of mutual protein concentrations in particular samples becomes possible, e.g., affected cell line versus negative control, patients with a specific disease versus healthy donors, etc. Two main quantification strategies are available at

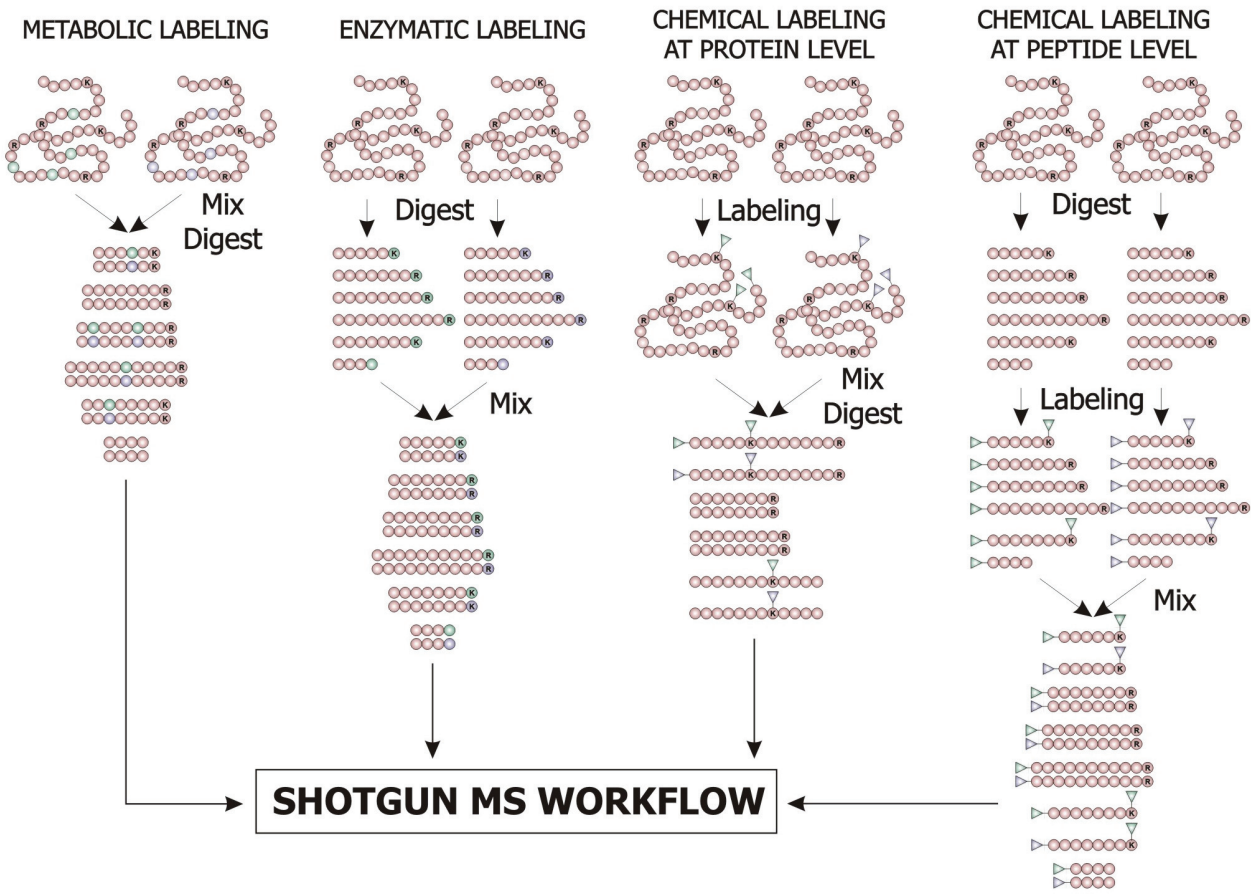

Fig. 4. Stable isotope labeling approaches. **Metabolic labeling**: Two cell cultures are grown in standard medium and a in medium containing heavy isotope labeled amino acids. After cultivation, cells are combined and are analyzed as a single sample. **Enzymatic labeling**: Two protein samples are digested by a sequence specific protease in either light (H_2O^{16}) or heavy (H_2O^{18}) water. Samples may be combined afterwards and processed as one. **Chemical labeling at protein level**: Proteins in two samples to be compared are labeled by ICAT reagents. After labeling, proteins are digested into peptides and combined. **Chemical labeling at peptide level**: Protein samples are digested separately into peptides. After digestion, each peptide sample is labeled by chemical reagents, which have identical chemical structure, but differ in stable isotope composition. After labeling, samples are combined and analyzed.

present: label-free quantification and quantification based on stable isotope labeling.

Label-free quantification

The label-free approach is based on comparison of MS signal intensities between individual experiments (Bondarenko *et al.* 2002). Semi-quantification is also possible to some extent by counting the number of peptides unambiguously identified (Ishihama *et al.* 2005). This method has several evident advantages and possible applications. The labeling step can be omitted, which both shortens and cheapens the experiment. The number of samples to be compared is virtually unlimited, which cannot be rivaled by any of the stable isotope based methods. Also, the spectral complexity is not increased, which could in turn lead to a higher number of identified proteins. Last but not least, label-free approaches are able to quantify throughout a much broader dynamic concentration range than stable isotope-based methods can. However, as different peptides ionize differently during individual experiments, their intensities may vary from run to run, making it rather difficult to correctly quantify them. Therefore, label-free methods are the least accurate, which is caused by the influence of both systematic and random errors during the experiment (Bantscheff *et al.* 2007). Nevertheless, techniques to overcome these shortcomings using bioinformatics and specialized software were suggested recently (Cox and Mann 2008). Thus, label-free quantitative proteomics particularly in combination with high resolution mass spectrometry (FT-ICR, Orbitrap) is regarded as a promising way to quantify large sets of samples even across multiple laboratories.

Labeling based on stable isotopes

Stable isotope strategies were introduced to deal

Table 3. Overview of stable isotope labeling methods in proteomics.

	SILAC	$^{16}O/^{18}O$	ICAT cICAT	Reductive alkylation	NBS	ICPL	TMT	iTRAQ
Labeling	Metabolic	Enzymatic	Chemical	Chemical	Chemical	Chemical	Chemical	Chemical
Labeling level	Proteins	Peptides	Proteins	Proteins Peptides	Peptides	Proteins Peptides	Peptides	Peptides
Target amino acid	L, R, K	C-terminus	C	N-terminus, K	W	N-terminus, K	N-terminus, K	N-terminus, K
Complexity reduction	No	No	Yes	No	Yes	No	No	No
Number of channels	2/3	2	2	2/3	2	2/3	2/6	4/8
Quantification mode	MS	MS	MS	MS	MS	MS	MSMS	MSMS

with the ionization variability of peptides and effect of errors during the workflow (Fig. 4). The samples to be compared can be mixed together and analyzed as a single one, whereas the combination of samples should be carried as soon as possible in the workflow. To distinguish the samples mixed during the analysis, they first need to be labeled with reagents containing stable isotopes, e.g. $^{12}C/^{13}C$, $^{14}N/^{15}N$, $^{16}O/^{18}O$ (Putz *et al.* 2005, Bantscheff *et al.* 2007). The proteins or peptides labeled with a substance of identical chemical structure containing stable isotopes will behave equally during all steps of the experiment since they have identical physicochemical (most importantly ionization and chromatographic) properties, but owing to a specific mass difference in their *m/z*, they can be simply recognized by a mass spectrometer. The quantification is then based on comparison of signal intensities, which differ by a specific molecular mass shift. Based on the nature of the sample, a broad range of quantification methods is at disposal. Stable isotopes may be incorporated into the samples metabolically, enzymatically, or by a chemical reaction (Table 3).

The Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) method is based on metabolic incorporation of amino acids containing stable isotopes into the protein sequence during cell culture cultivation in a medium containing either light or heavy forms of particular amino acids, e.g. leucine or arginine (Ong *et al.* 2002, Blagoev *et al.* 2004). Owing to the metabolic nature of the labeling, the SILAC method cannot be directly employed in proteomic analyses aimed at plasma biomarker discovery. However, SILAC recently became a basis for a novel combined strategy for biomarker identification called the Stable Isotope Labeled Proteome (SILAP) method (Shah *et al.* 2009, Yu *et al.* 2009). Briefly, a cell model of the studied disease, i.e., pancreatic cancer cell line, is grown in a heavy form of the SILAC cell culture medium. Labeled proteins from these cells that are secreted into the medium are collected and subsequently combined with human plasma samples from patients suffering from pancreatic cancer and with plasma from healthy controls. Due to the heavy isotope labeling, proteins originating from the cell line are recognized in the mass spectrum as they differ by a specific mass shift from the same protein in plasma. The ratios of secretome versus control plasma and secretome versus diseased plasma are then compared, and proteins with altered ratios may than be considered as candidate markers, suitable for subsequent validation.

The next possible point in the shotgun proteomics workflow suitable for labeling is the enzymatic digestion of proteins into peptides, since certain proteases, e.g. trypsin, Glu-C, and Lys-C, catalyze exchange of two oxygen atoms at the C-termini of the peptides by two oxygen atoms coming from solvent water during the reaction (Schnolzer *et al.* 1996). When two protein samples to be compared are digested in $H_2{}^{16}O$ and $H_2{}^{18}O$ separately, the resulting peptides differ by 4 daltons (Da), which is sufficient to recognize peptide pairs properly in the mass spectrum (Heller *et al.* 2003, Havlis and Shevchenko 2004).

The incorporation of stable isotopes by a

chemical reaction represents the largest group of quantification methods. The very first chemical quantification method was the Isotope-Coded Affinity Tags (ICAT) approach, which is based on labeling cysteine-containing peptides via their thiol groups. Light and heavy ICAT labels also contain biotin; therefore the labeled peptides can be isolated using avidin. Due to the fact that approximately one quarter of all tryptic peptides contains cysteine, the enrichment results in significant reduction of the sample complexity (Liu *et al.* 2005). But as majority of proteins contain at least one cysteine in their structure, the information on the originating protein is not lost (Gygi *et al.* 1999). Cystein can be targeted also by other chemistries, such as those based on acrylamide reaction (Faca *et al.* 2008).

The majority of chemical quantification methods incorporate stable isotopes into the peptides using a reaction of $-NH_2$ groups with succinimide derivates. To illustrate, the Isotope-Coded Protein Label Triple (ICPL) method uses N-nicotinoyloxysuccinimide and offers up to three quantification channels (Schmidt *et al.* 2005). NH_2 groups may be also tagged by more stable and even less expensive chemistry based on reductive alkylation using formaldehyde (Boersema *et al.* 2008). In theory, $-NH_2$ groups-targeted labeling covers all the peptides resulting from a protein digest. A significant bottleneck of these techniques emerges during labeling at protein level because the altered side chain of lysine is not recognized by trypsin and thus incomplete cleavage occurs, resulting in fewer and larger peptides. If one wishes to preserve trypsin cleavage rather than select another protease, other functional groups must be tagged at protein level. In this case, however, peptides lacking the target group do not carry quantitative information. On the other hand, by introducing the isolation/enrichment step only or more frequently peptides carrying the tag may be analyzed – lowering the sample complexity as described in ICAT (Gygi *et al.* 1999) or NBS method (Matsuo *et al.* 2009).

Most of the labeling techniques are based on quantification at MS level, where the MS spectra are searched for signals differing by a specific *m/z* shift. The relative concentration of a given peptide is then obtained by comparing the intensities of these corresponding signals.

MS-based quantification techniques enable analysis of a limited number of samples simultaneously, whereas MSMS-based isobaric techniques offer a much higher number of possible quantification channels. The isobaric labels used in these techniques are composed of a reactive group, a reported group, and a balancer group. The sum of molecular weight of these three parts is constant, therefore a labeled peptide is observed as a single peak in MS mode. But as the individual reporter groups differ in molecular weight, the MSMS fragments originating from these reporter groups are observable as distinct peaks. Relative peptide concentration may be acquired by comparing the MSMS signal intensities of these reporter groups.

The Tandem Mass Tags were the first published isobaric technique (Thompson *et al.* 2003). The Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) method, however, has gained greater popularity, as it enables an analysis of up to four samples simultaneously (Ross *et al.* 2004), the newest version even up to eight different samples in one experiment (Pierce *et al.* 2007). During the fragmentation in MSMS mode, the reporter group is released from the modified peptides and can be observed in the mass spectrum as peaks 114.1, 115.1, 116.1 or 117.1 (Fig. 5). Therefore, a tandem mass spectrometer capable of detecting MSMS fragments in low *m/z* range is mandatory. The acquired MSMS spectra are used both for peptide identification and for quantification, where the reporter group signals are used to calculate relative peptide concentrations in particular samples and the remaining fragments originating from the peptide backbone are used for peptide identification. To obtain the reporter ion signal in the spectrum, the reporter group has to be cleaved from the peptide properly. Our data show that the cleavage efficacy varies based on the peptide structure. Nevertheless, because the character of the bond is identical in all four or eight tags, respectively, the cleavage efficacy from a particular peptide is also supposed to remain constant.

Targeted shotgun proteomics

Until recently, MS has been used almost exclusively for the identification of new potential biomarkers, whereas the verification and validation steps were carried out by antibody-based techniques. Recently, a paradigm shift has been apparent, as targeted tandem mass spectrometry also known as Multiple Reaction Monitoring (LC-MRM-MSMS) is increasingly being applied into both verification and validation phases (Lange *et al.* 2008). Interestingly, this targeted approach has been also implemented into multistage strategies for biomarker identification, due to excellent sensitivity and potential to precisely quantify target molecules in complex samples. This is carried out by detecting signature peptides, which

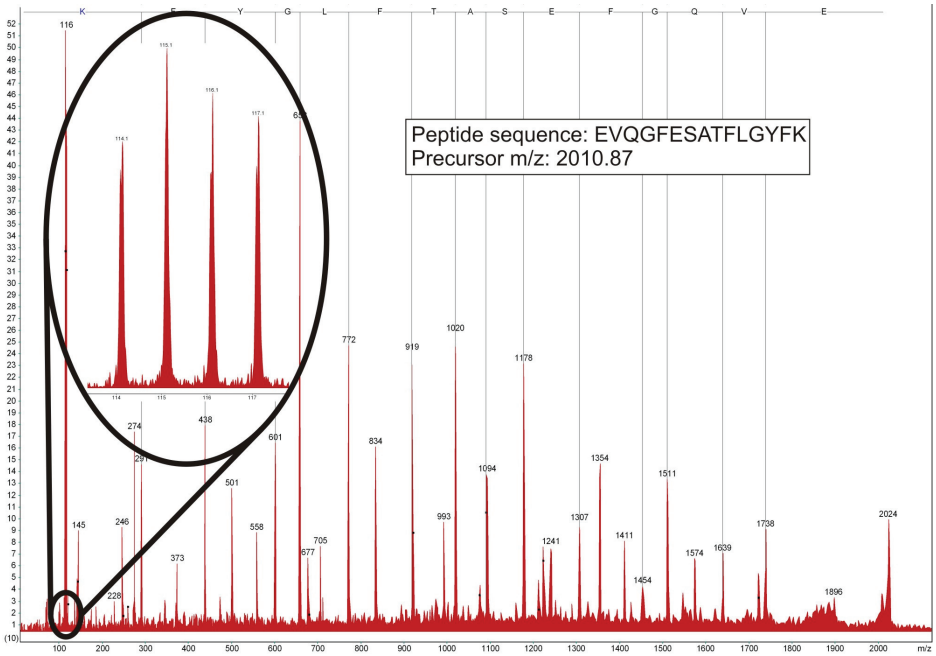

Fig. 5. Representative iTRAQ MSMS spectrum: A peptide of m/z 2010.87 was selected from a MS spectrum for fragmentation analysis. The resulting MSMS spectrum is shown. iTRAQ quantitation information can be read in the low m/z region, as shown in the magnified view. The intensity of each of the four peaks (114.1, 115.1, 116.1 and 117.1) reflects relative concentration of the given peptide in individual four samples which are to be compared. The sequence EVQGFESATFLGYFK was successfully assigned to the MSMS spectrum, resulting in unambiguous identification of Isoform 2 of Gelsolin precursor.

are unique for a given protein, by LC-MRM-MSMS (Anderson and Hunter 2006, Kitteringham *et al.* 2009). Triple quadrupole (QqQ) or hybrid quadrupole-linear ion traps (QTRAP) mass spectrometers used for this purpose are set to select only a specific precursor peptide in the first quadrupole (Q1), which is then fragmented in the collision cell (Q2) and a specific fragment is selected in the third quadrupole (Q3) (Fig. 6). As this cycle takes only a few milliseconds, tens to hundreds of different peptides may be detected and quantified in a targeted manner during a single LC run. More importantly, the detection limit for peptides in this configuration in enhanced by up to 100-fold as opposed to unbiased MS analysis (Keshishian *et al.* 2007). To reliably confirm the identity of monitored peptide, a full MSMS scan upon detecting a defined MRM transition should be acquired (Unwin *et al.* 2009). By this means, the peptide is precisely quantified by the respective chromatographic peak and confirmed by sequence acquisition from the MSMS spectrum (Fig. 7). The actual quantification is carried out by plotting the intensity of Q3 fragment ions on time axis, which results in a chromatographic peak. The most accurate way of quantifying among more samples is realized by introducing a synthetic internal standard peptide, containing a heavy amino acid, into the analyzed sample. As already described in previous chapters, these labeled peptides follow their natural counterparts during all steps of analysis, but owing to a specific mass difference, they can be easily distinguished by the mass spectrometer. The peak area of internal standard peptide, where the precise concentration is know, is compared to the peak area of peptide originating from analyzed sample and finally, absolute concentration may be calculated.

Table 4. Individual phases of a biomarker discovery pipeline (Rifai *et al.* 2006).

Phase I	Exploratory studies to identify candidate marker molecules
Phase II	Qualification – confirmation of differential abundance in samples
Phase III	Verification – assess specificity of candidate molecules
Phase IV	Validation and clinical assay development – large scale studies

The role of proteomics in biomarker candidates verification

Regardless of the method used as the first step of the biomarker discovery process, the resulting candidate markers need to be further intensively proved and tested if they are to become clinically used biomarkers. This is a multistage process and can be regarded as an analogy to the drug discovery pipeline. Starting with a large group of marker candidates, the funnel-like process eliminates low-sensitive and low-specific markers, resulting in a few final candidate molecules. Proportionally to the candidate marker count reduction, the number of tested samples grows steeply along with the project costs (Rifai *et al.* 2006) (Table 4).

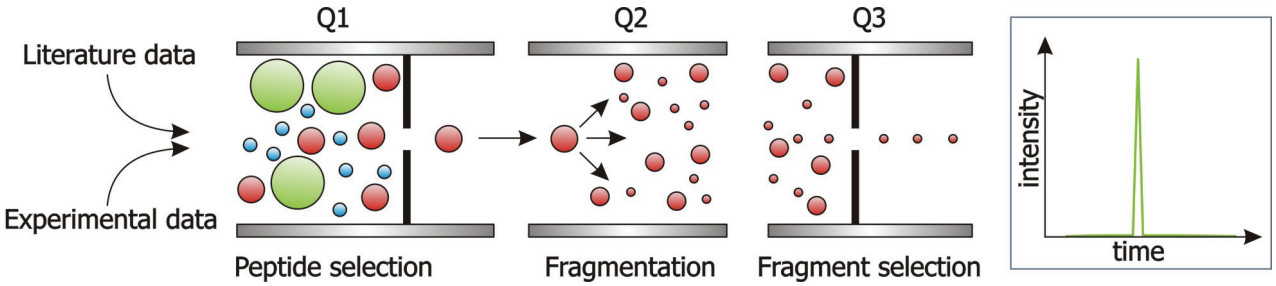

Fig. 6. Multiple Reaction Monitoring scheme. Mass spectrometers used for MRM are set to select only a specific precursor peptide in the first quadrupole (Q1), which is then fragmented in the collision cell (Q2), a specific fragment is selected in the third quadrupole (Q3) and detected. The intensity of the Q3 fragment is then plotted in time, which results in a chromatographic peak correlating with peptide amount in the sample.

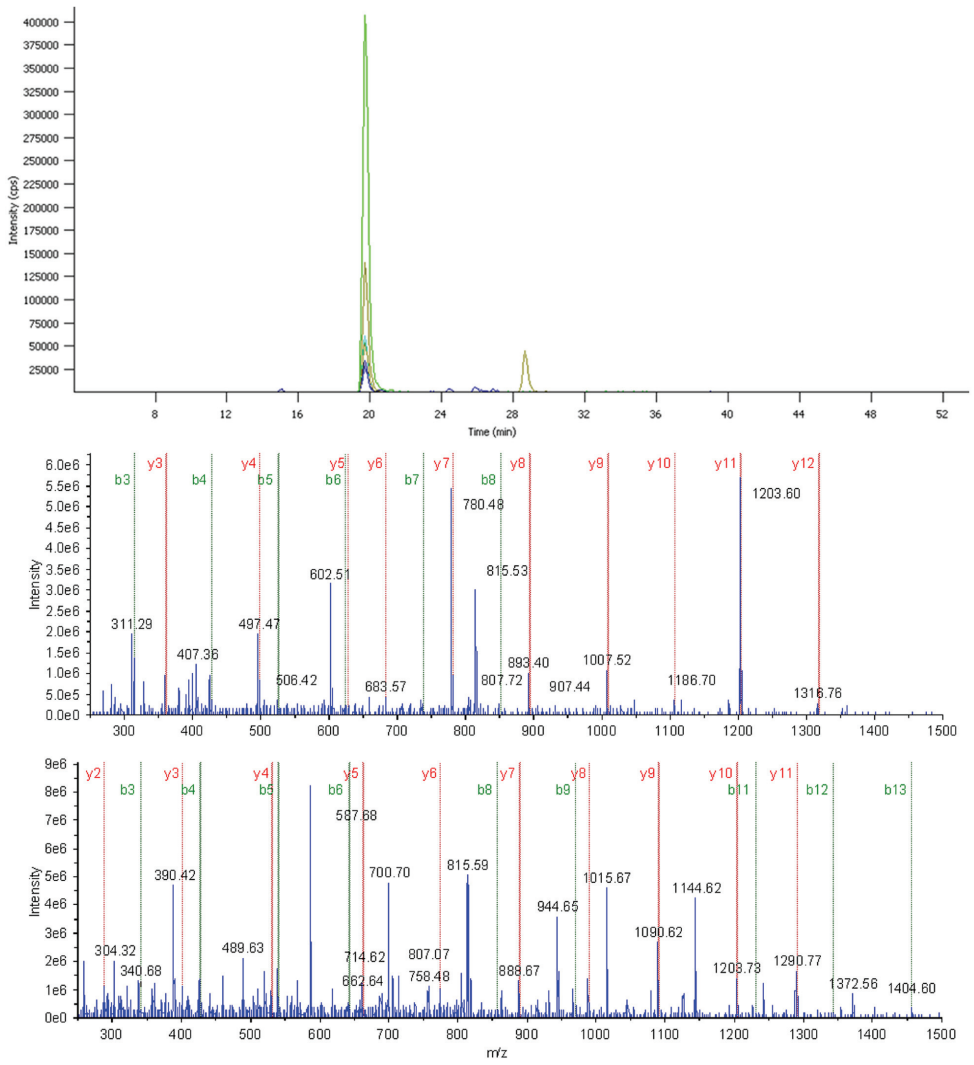

Fig. 7. MRM triggered MSMS. A confident MRM assay should be validated by confirming the identity of the chromatographic MRM peaks by additional acquisition of a MSMS spectrum. The MRM trace shows two chromatographic peaks, each eluting at a different time point. By acquiring a MRM-triggered MSMS spectrum, the targeted earlier eluting signature peptide is unambiguously identified (upper MSMS spectrum), whereas the second peak (lower MSMS spectrum) was proved to originate from a different protein.

After the first phase of the pipeline, the discovery phase, the resulting candidates need to be further proved in the qualification phase, in order to confirm their differential abundance in the tested samples. At this point, the unbiased nature of the analysis changes into a targeted one. New and unproven candidates are analyzed in a targeted manner and precisely quantified in a statistically viable number of serum or plasma samples. Unfortunately, antibodies against these newly discovered candidates are frequently unavailable, and substitutes for antibody-based detection assays (i.e., Western blotting or ELISA) have been sought in proteomics methodologies. Therefore, the method of choice in this phase is LC-MRM-MSMS (Anderson and

Hunter 2006). As the sample preparation and processing is much less extensive than in the discovery phase, the MRM sensitivity is limited by the sample complexity. The limit of quantification in undepleted plasma may reach low μg/ml level (Addona *et al.* 2009). To quantify in the ng/ml range, depletion of ~10 most abundant plasma proteins is required (Keshishian *et al.* 2007). A possible way to further increase the performance of MRM is its coupling to immunoaffinity peptide enrichment (Anderson *et al.* 2004a, Hoofnagle *et al.* 2008), which enhances both sensitivity and specificity, thus allowing analysis in complex matrices with little or no fractionation. However, it requires a specific anti-peptide antibody to be developed against each analyzed peptide. Other antibody-based approaches are unsuitable at this point, due to their low throughput, e.g., Western blotting or high development costs typical for immunoassays.

In the verification phase the specificity of candidates is addressed. The primary objective of verification is to screen potential biomarkers to ensure that only the highest-quality candidates from the discovery phase are taken forward into pre-clinical validation. This requires a larger number of tested samples, which increases approximately by an order. So as to maintain a moderate throughput, the initial candidate list has to be reduced to a few dozens. Immunoassays should be introduced at this point. The lack of high-quality antibodies, however, hinders the fast development of antibody-based assays, as highly specific antibodies are not available for most novel biomarker candidates. Therefore, LC-MRM-MSMS presents a compelling alternative to immunoassays, as it allows a moderate number of candidates to be targeted at a relatively high throughput, without a need of an immunoassay development.

The final phase of the biomarker discovery process, the validation phase, requires a clinical assay to be developed and extensively tested on thousands of clinical samples. A platform change is also required, as MS-based approaches are currently neither able to fulfill the required combination of high throughput and precision, nor are they widely available and accepted by the FDA. Therefore, the development of a suitable antibody-based assay is mandatory (Kingsmore 2006). To meet the required sensitivity, RIA or ELISA are the methods of choice.

Conflict of Interest

There is no conflict of interest.

Acknowledgements

The authors thank Martin Hubalek for critical comments and useful remarks. This article was supported by grant number NR9253-3, provided by Ministry of Health of the Czech Republic.

Abreviations

2D-PAGE – two-dimensional polyacrylamide gel electrophoresis, amu – atomic mass unit, DIGE – differential gel electrophoresis, ELISA – enzyme-linked immunosorbent assay, ESI – electrospray ionization, FDA – Food and Drug Administration, FT-ICR – Fourier-transform ion cyclotron resonance, HPLC – high performance liquid chromatography, HUPO – Human Proteome Organisation, ICAT – isotope-coded affinity tags, ICPL – isotope-coded protein label, IEF – isoelectric focusing, IPG – immobilized pH gradient, IT – ion trap, iTRAQ – isobaric tags for relative and absolute quantitation, LC – liquid chromatography, *m/z* – mass to charge ratio, MALDI – matrix-assisted laser desorption/ionization, MARS – Multiple Affinity Removal System, MRM – multiple reaction monitoring, MS – mass spectrometry, MSMS – tandem mass spectrometry, MW – molecular weight, NBS – 2-nitrobenzenesulfenyl, pI – isoelectric point, PMF – peptide mass fingerprinting, Q – quadrupole, QqQ – triple quadrupole, Q-TRAP – quadrupole-ion trap, RP – reversed-phase, SCX – strong cation exchange, SDS – sodium dodecyl sulfate, SELDI – surface-enhanced laser desorption/ionization, SILAC – stable isotope labeling with amino acids in cell culture, SILAP – stable isotope labeled proteome, TOF – time-of-flight

References

ADDONA TA, ABBATIELLO SE, SCHILLING B, SKATES SJ, MANI DR, BUNK DM, SPIEGELMAN CH, ZIMMERMAN LJ, HAM AJ, KESHISHIAN H, HALL SC, ALLEN S, BLACKMAN RK, BORCHERS CH, BUCK C, CARDASIS HL, CUSACK MP, DODDER NG, GIBSON BW, HELD JM, HILTKE T, JACKSON A, JOHANSEN EB, KINSINGER CR, LI J, MESRI M, NEUBERT TA, NILES RK, PULSIPHER TC, RANSOHOFF D, RODRIGUEZ H, RUDNICK PA, SMITH D, TABB DL, TEGELER TJ, VARIYATH AM, VEGA-MONTOTO LJ, WAHLANDER A, WALDEMARSON S, WANG M, WHITEAKER JR, ZHAO L, ANDERSON NL, FISHER SJ, LIEBLER DC, PAULOVICH AG, REGNIER FE, TEMPST P, CARR SA: Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **27**: 633-641, 2009.

ANDERSON L: Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J Physiol Lond* **563**: 23-60, 2005.

ANDERSON L, HUNTER CL: Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* **5**: 573-588, 2006.

ANDERSON NL, ANDERSON NG: The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**: 845-867, 2002.

ANDERSON NL, ANDERSON NG, HAINES LR, HARDIE DB, OLAFSON RW, PEARSON TW: Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res* **3**: 235-244, 2004a.

ANDERSON NL, POLANSKI M, PIEPER R, GATLIN T, TIRUMALAI RS, CONRADS TP, VEENSTRA TD, ADKINS JN, POUNDS JG, FAGAN R, LOBLEY A: The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics* **3**: 311-326, 2004b.

ANDRIOLE GL, CRAWFORD ED, GRUBB RL 3RD, BUYS SS, CHIA D, CHURCH TR, FOUAD MN, GELMANN EP, KVALE PA, REDING DJ, WEISSFELD JL, YOKOCHI LA, O'BRIEN B, CLAPP JD, RATHMELL JM, RILEY TL, HAYES RB, KRAMER BS, IZMIRLIAN G, MILLER AB, PINSKY PF, PROROK PC, GOHAGAN JK, BERG CD: Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**: 1310-1319, 2009.

ATKINSON AJ, COLBURN WA, DEGRUTTOLA VG: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69**: 89-95, 2001.

BANKS RE: Preanalytical influences in clinical proteomic studies: raising awareness of fundamental issues in sample banking. *Clin Chem* **54**: 6-7, 2008.

BANTSCHEFF M, SCHIRLE M, SWEETMAN G, RICK J, KUSTER B: Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**: 1017-1031, 2007.

BAZHIN AV, DALKE C, WILLNER N, ABSCHUTZ O, WILDBERGER HG, PHILIPPOV PP, DUMMER R, GRAW J, DE ANGELIS MH, SCHADENDORF D, UMANSKY V, EICHMULLER SB: Cancer-retina antigens as potential paraneoplastic antigens in melanoma-associated retinopathy. *Int J Cancer* **124**: 140-149, 2009.

BECK LH JR, BONEGIO RG, LAMBEAU G, BECK DM, POWELL DW, CUMMINS TD, KLEIN JB, SALANT DJ: M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy. *N Engl J Med* **361**: 11-21, 2009.

BERGGREN K, CHERNOKALSKAYA E, STEINBERG TH, KEMPER C, LOPEZ MF, DIWU Z, HAUGLAND RP, PATTON WF: Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* **21**: 2509-2521, 2000.

BERTH M, MOSER FM, KOLBE M, BERNHARDT J: The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl Microbiol Biotechnol* **76**: 1223-1243, 2007.

BIJIAN K, MLYNAREK AM, BALYS RL, JIE S, XU Y, HIER MP, BLACK MJ, DI FALCO MR, LABOISSIERE S, ALAOUI-JAMALI MA: Serum proteomic approach for the identification of serum biomarkers contributed by oral squamous cell carcinoma and host tissue microenvironment. *J Proteome Res* **8**: 2173-2185, 2009.

BJELLQVIST B, EK K, RIGHETTI PG, GIANAZZA E, GORG A, WESTERMEIER R, POSTEL W: Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**: 317-339, 1982.

BLAGOEV B, ONG SE, KRATCHMAROVA I, MANN M: Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22**: 1139-1145, 2004.

BODNAR WM, BLACKBURN RK, KRISE JM, MOSELEY MA: Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom* **14**: 971-979, 2003.

BOERSEMA PJ, AYE TT, VAN VEEN TA, HECK AJ, MOHAMMED S: Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **8**: 4624-4632, 2008.

BONDARENKO PV, CHELIUS D, SHALER TA: Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* **74**: 4741-4749, 2002.

BROOKS SA, CARTER TM, ROYLE L, HARVEY DJ, FRY SA, KINCH C, DWEK RA, RUDD PM: Altered glycosylation of proteins in cancer: what is the potential for new anti-tumour strategies. *Anticancer Agents Med Chem* **8**: 2-21, 2008.

BRUEGEL M, PLANERT M, BAUMANN S, FOCKE A, BERGH FT, LEICHTLE A, CEGLAREK U, THIERY J, FIEDLER GM: Standardized peptidome profiling of human cerebrospinal fluid by magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Proteomics* **72**: 608-615, 2009.

CALLESEN AK, MADSEN JS, VACH W, KRUSE TA, MOGENSEN O, JENSEN ON: Serum protein profiling by solid phase extraction and mass spectrometry: a future diagnostics tool? *Proteomics* **9**: 1428-1441, 2009.

CANDIANO G, BRUSCHI M, MUSANTE L, SANTUCCI L, GHIGGERI GM, CARNEMOLLA B, ORECCHIA P, ZARDI L, RIGHETTI PG: Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**: 1327-1333, 2004.

CARRETTE O, BURKHARD PR, SANCHEZ JC, HOCHSTRASSER DF: State-of-the-art two-dimensional gel electrophoresis: a key tool of proteomics research. *Nat Protoc* **1**: 812-823, 2006.

CHEN HS, REJTAR T, ANDREEV V, MOSKOVETS E, KARGER BL: High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* **77**: 2323-2331, 2005.

CHEVALLET M, LUCHE S, DIEMER H, STRUB JM, VAN DORSSELAER A, RABILLOUD T: Sweet silver: A formaldehyde-free silver staining using aldoses as developing agents, with enhanced compatibility with mass spectrometry. *Proteomics* **8**: 4853-4861, 2008.

CHROMY BA, GONZALES AD, PERKINS J, CHOI MW, CORZETT MH, CHANG BC, CORZETT CH, MCCUTCHEN-MALONEY SL: Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J Proteome Res* **3**: 1120-1127, 2004.

CLAUSER KR, BAKER P, BURLINGAME AL: Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **71**: 2871-2882, 1999.

COLLINS FS, LANDER ES, ROGERS J, WATERSTON RH: Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931-945, 2004.

COMEGYS MM, LIN SH, RAND D, BRITT D, FLANAGAN D, CALLANAN H, BRILLIANT K, HIXSON DC: Two variable regions in carcinoembryonic antigen-related cell adhesion molecule1 N-terminal domains located in or next to monoclonal antibody and adhesion epitopes show evidence of recombination in rat but not in human. *J Biol Chem* **279**: 35063-35078, 2004.

COX J, MANN M: High peptide identification rates and proteome-wide quantitation via novel computational strategies. *Nat Biotechnol* **26**: 1367-1372, 2008.

DATTELBAUM AM, IYER S: Surface-assisted laser desorption/ionization mass spectrometry. *Expert Rev Proteomics* **3**: 153-161, 2006.

DESOUZA LV, ROMASCHIN AD, COLGAN TJ, SIU KW: Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. *Anal Chem* **81**: 3462-3470, 2009.

DESOUZA LV, TAYLOR AM, LI W, MINKOFF MS, ROMASCHIN AD, COLGAN TJ, SIU KW: Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J Proteome Res* **7**: 3525-3534, 2008.

DOMON B, AEBERSOLD R: Mass spectrometry and protein analysis. *Science* **312**: 212-217, 2006.

ECHAN LA, TANG HY, ALI-KHAN N, LEE K, SPEICHER DW: Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* **5**: 3292-3303, 2005.

ESSADER AS, CARGILE BJ, BUNDY JL, STEPHENSON JL JR: A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **5**: 24-34, 2005.

ETZIONI R, URBAN N, RAMSEY S, MCINTOSH M, SCHWARTZ S, REID B, RADICH J, ANDERSON G, HARTWELL L: The case for early detection. *Nat Rev Cancer* **3**: 243-252, 2003.

FACA VM, SONG KS, WANG H, ZHANG Q, KRASNOSELSKY AL, NEWCOMB LF, PLENTZ RR, GURUMURTHY S, REDSTON MS, PITTERI SJ, PEREIRA-FACA SR, IRETON RC, KATAYAMA H, GLUKHOVA V, PHANSTIEL D, BRENNER DE, ANDERSON MA, MISEK D, SCHOLLER N, URBAN ND, BARNETT MJ, EDELSTEIN C, GOODMAN GE, THORNQUIST MD, MCINTOSH MW, DEPINHO RA, BARDEESY N, HANASH SM: A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med* **5**: e123, 2008.

FIEDLER GM, LEICHTLE AB, KASE J, BAUMANN S, CEGLAREK U, FELIX K, CONRAD T, WITZIGMANN H, WEIMANN A, SCHUTTE C, HAUSS J, BUCHLER M, THIERY J: Serum peptidome profiling revealed platelet factor 4 as a potential discriminating Peptide associated with pancreatic cancer. *Clin Cancer Res* **15**: 3812-3819, 2009.

FINDEISEN P, ZAPATKA M, PECCERELLA T, MATZK H, NEUMAIER M, SCHADENDORF D, UGUREL S: Serum amyloid A as a prognostic marker in melanoma identified by proteomic profiling. *J Clin Oncol* **27**: 2199-2208, 2009.

GILAR M, OLIVOVA P, CHAKRABORTY AB, JAWORSKI A, GEROMANOS SJ, GEBLER JC: Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum. *Electrophoresis* **30**: 1157-1167, 2009.

GOLDMAN R, RESSOM HW, ABDEL-HAMID M, GOLDMAN L, WANG A, VARGHESE RS, AN Y, LOFFREDO CA, DRAKE SK, EISSA SA, GOUDA I, EZZAT S, MOISEIWITSCH FS: Candidate markers for the detection of hepatocellular carcinoma in low-molecular weight fraction of serum. *Carcinogenesis* **28**: 2149-2153, 2007.

GOLDMAN R, RESSOM HW, VARGHESE RS, GOLDMAN L, BASCUG G, LOFFREDO CA, ABDEL-HAMID M, GOUDA I, EZZAT S, KYSELOVA Z, MECHREF Y, NOVOTNY MV: Detection of hepatocellular carcinoma using glycomic analysis. *Clin Cancer Res* **15**: 1808-1813, 2009.

GORG A, WEISS W, DUNN MJ: Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**: 3665-3685, 2004.

GOVORUKHINA NI, DE VRIES M, REIJMERS TH, HORVATOVICH P, VAN DER ZEE AG, BISCHOFF R: Influence of clotting time on the protein composition of serum samples based on LC-MS data. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**: 1281-1291, 2009.

GRAMOLINI AO, KISLINGER T, ALIKHANI-KOOPAEI R, FONG V, THOMPSON NJ, ISSERLIN R, SHARMA P, OUDIT GY, TRIVIERI MG, FAGAN A, KANNAN A, HIGGINS DG, HUEDIG H, HESS G, ARAB S, SEIDMAN JG, SEIDMAN CE, FREY B, PERRY M, BACKX PH, LIU PP, MACLENNAN DH, EMILI A: Comparative proteomics profiling of a phospholamban mutant mouse model of dilated cardiomyopathy reveals progressive intracellular stress responses. *Mol Cell Proteomics* **7**: 519-533, 2008.

GUO Y, FU Z, VAN EYK JE: A proteomic primer for the clinician. *Proc Am Thorac Soc* **4**: 9-17, 2007.

GYGI SP, RIST B, GERBER SA, TURECEK F, GELB MH, AEBERSOLD R: Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**: 994-999, 1999.

HANASH SM, PITTERI SJ, FACA VM: Mining the plasma proteome for cancer biomarkers. *Nature* **452**: 571-579, 2008.

HASHIGUCHI T, TANAKA K, LEE LJ, SASAKI K, NATSUGOE S, KAWAHARA KI, ARIMURA K, MARUYAMA I: Diagnostic value of serum peptidome analyses for protease activated pathological conditions beyond cancer diagnosis. *Med Hypotheses* **73**: 760-763, 2009.

HAVLIS J, SHEVCHENKO A: Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Anal Chem* **76**: 3029-3036, 2004.

HELLER M, MATTOU H, MENZEL C, YAO X: Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J Am Soc Mass Spectrom* **14**: 704-718, 2003.

HENZEL WJ, WATANABE C, STULTS JT: Protein identification: the origins of peptide mass fingerprinting. *J Am Soc Mass Spectrom* **14**: 931-942, 2003.

HOFFMAN SA, JOO WA, ECHAN LA, SPEICHER DW: Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**: 43-52, 2007.

HONGSACHART P, HUANG-LIU R, SINCHAIKUL S, PAN FM, PHUTRAKUL S, CHUANG YM, YU CJ, CHEN ST: Glycoproteomic analysis of WGA-bound glycoprotein biomarkers in sera from patients with lung adenocarcinoma. *Electrophoresis* **30**: 1206-1220, 2009.

HOOFNAGLE AN, BECKER JO, WENER MH, HEINECKE JW: Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin Chem* **54**: 1796-1804, 2008.

HU S, LOO JA, WONG DT: Human body fluid proteome analysis. *Proteomics* **6**: 6326-6353, 2006.

HUANG HL, STASYK T, MORANDELL S, MOGG M, SCHREIBER M, FEUERSTEIN I, HUCK CW, STECHER G, BONN GK, HUBER LA: Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. *Electrophoresis* **26**: 2843-2849, 2005a.

HUANG L, FANG X: Immunoaffinity fractionation of plasma proteins by chicken IgY antibodies. *Methods Mol Biol* **425**: 41-51, 2008.

HUANG L, HARVIE G, FEITELSON JS, GRAMATIKOFF K, HEROLD DA, ALLEN DL, AMUNNGAMA R, HAGLER RA, PISANO MR, ZHANG WW, FANG X: Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. *Proteomics* **5**: 3314-3328, 2005b.

ISHIHAMA Y, ODA Y, TABATA T, SATO T, NAGASU T, RAPPSILBER J, MANN M: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**: 1265-1272, 2005.

KESHISHIAN H, ADDONA T, BURGESS M, KUHN E, CARR SA: Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **6**: 2212-2229, 2007.

KESHISHIAN H, ADDONA T, BURGESS M, MANI DR, SHI X, KUHN E, SABATINE MS, GERSZTEN RE, CARR SA: Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **8**: 2339-2349, 2009.

KIM HJ, KANG HJ, LEE H, LEE ST, YU MH, KIM H, LEE C: Identification of S100A8 and S100A9 as serological markers for colorectal cancer. *J Proteome Res* **8**: 1368-1379, 2009.

KINGSMORE SF: Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat Rev Drug Discov* **5**: 310-320, 2006.

KITTERINGHAM NR, JENKINS RE, LANE CS, ELLIOTT VL, PARK BK: Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**: 1229-1239, 2009.

KUI WONG N, EASTON RL, PANICO M, SUTTON-SMITH M, MORRISON JC, LATTANZIO FA, MORRIS HR, CLARK GF, DELL A, PATANKAR MS: Characterization of the oligosaccharides associated with the human ovarian tumor marker CA125. *J Biol Chem* **278**: 28619-28634, 2003.

KULASINGAM V, DIAMANDIS EP: Tissue culture-based breast cancer biomarker discovery platform. *Int J Cancer* **123**: 2007-2012, 2008.

KYSELOVA Z, MECHREF Y, KANG P, GOETZ JA, DOBROLECKI LE, SLEDGE GW, SCHNAPER L, HICKEY RJ, MALKAS LH, NOVOTNY MV: Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem* **54**: 1166-1175, 2008.

LAEMMLI UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685, 1970.

LANGE V, PICOTTI P, DOMON B, AEBERSOLD R: Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4**: 222, 2008.

LAU TK, LEUNG TN: Genetic screening and diagnosis. *Curr Opin Obstet Gynecol* **17**: 163-169, 2005.

LIU T, QIAN WJ, CHEN WN, JACOBS JM, MOORE RJ, ANDERSON DJ, GRITSENKO MA, MONROE ME, THRALL BD, CAMP DG 2ND, SMITH RD: Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: the human mammary epithelial cell proteome. *Proteomics* **5**: 1263-1273, 2005.

LIU T, QIAN WJ, MOTTAZ HM, GRITSENKO MA, NORBECK AD, MOORE RJ, PURVINE SO, CAMP DG 2ND, SMITH RD: Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol Cell Proteomics* **5**: 2167-2174, 2006.

LOOI KS, NAKAYASU ES, DIAZ RA, TAN EM, ALMEIDA IC, ZHANG JY: Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma. *J Proteome Res* **7**: 4004-4012, 2008.

LUDWIG JA, WEINSTEIN JN: Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* **5**: 845-856, 2005.

MACKINTOSH JA, CHOI HY, BAE SH, VEAL DA, BELL PJ, FERRARI BC, VAN DYK DD, VERRILLS NM, PAIK YK, KARUSO P: A fluorescent natural product for ultra sensitive detection of proteins in one-dimensional and two-dimensional gel electrophoresis. *Proteomics* **3**: 2273-2288, 2003.

MANN M, WILM M: Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**: 4390-4399, 1994.

MATSUO EI, WATANABE M, KUYAMA H, NISHIMURA O: A new strategy for protein biomarker discovery utilizing 2-nitrobenzenesulfenyl (NBS) reagent and its applications to clinical samples. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**: 2607-2614, 2009.

MECHREF Y, MADERA M, NOVOTNY MV: Glycoprotein enrichment through lectin affinity techniques. *Methods Mol Biol* **424**: 373-396, 2008.

MILLER I, CRAWFORD J, GIANAZZA E: Protein stains for proteomic applications: which, when, why? *Proteomics* **6**: 5385-5408, 2006.

MITULOVIC G, MECHTLER K: HPLC techniques for proteomics analysis--a short overview of latest developments. *Brief Funct Genomic Proteomic* **5**: 249-260, 2006.

NESVIZHSKII AI: Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol* **367**: 87-120, 2006.

NEUHOFF V, AROLD N, TAUBE D, EHRHARDT W: Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**: 255-262, 1988.

O'FARRELL PH: High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007-4021, 1975.

OMENN GS, STATES DJ, ADAMSKI M, BLACKWELL TW, MENON R, HERMJAKOB H, APWEILER R, HAAB BB, SIMPSON RJ, EDDES JS, KAPP EA, MORITZ RL, CHAN DW, RAI AJ, ADMON A, AEBERSOLD R, ENG J, HANCOCK WS, HEFTA SA, MEYER H, PAIK YK, YOO JS, PING P, POUNDS J, ADKINS J, QIAN X, WANG R, WASINGER V, WU CY, ZHAO X, ZENG R, ARCHAKOV A, TSUGITA A, BEER I, PANDEY A, PISANO M, ANDREWS P, TAMMEN H, SPEICHER DW, HANASH SM: Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**: 3226-3245, 2005.

ONG SE, BLAGOEV B, KRATCHMAROVA I, KRISTENSEN DB, STEEN H, PANDEY A, MANN M: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376-386, 2002.

PAPPIN DJ, HOJRUP P, BLEASBY AJ: Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**: 327-332, 1993.

PIERCE A, UNWIN RD, EVANS CA, GRIFFITHS S, CARNEY L, ZHANG L, JAWORSKA E, LEE CF, BLINCO D, OKONIEWSKI MJ, MILLER CJ, BITTON DA, SPOONCER E, WHETTON AD: Eight-channel iTRAQ enables comparison of the activity of 6 leukaemogenic tyrosine kinases. *Mol Cell Proteomics* **7**: 853-863, 2008.

POON TC: Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. *Expert Rev Proteomics* **4**: 51-65, 2007.

PUTZ S, REINDERS J, REINDERS Y, SICKMANN A: Mass spectrometry-based peptide quantification: applications and limitations. *Expert Rev Proteomics* **2**: 381-392, 2005.

QUINTANA LF, CAMPISTOL JM, ALCOLEA MP, BANON-MANEUS E, SOL-GONZALEZ A, CUTILLAS PR: Application of label-free quantitative peptidomics for the identification of urinary biomarkers of kidney chronic allograft dysfunction. *Mol Cell Proteomics* **8**: 1658-1673, 2009.

RABILLOUD T, VUILLARD L, GILLY C, LAWRENCE JJ: Silver-staining of proteins in polyacrylamide gels: a general overview. *Cell Mol Biol (Noisy-le-grand)* **40**: 57-75, 1994.

RAI AJ, GELFAND CA, HAYWOOD BC, WARUNEK DJ, YI J, SCHUCHARD MD, MEHIGH RJ, COCKRILL SL, SCOTT GB, TAMMEN H, SCHULZ-KNAPPE P, SPEICHER DW, VITZTHUM F, HAAB BB, SIEST G, CHAN DW: HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**: 3262-3277, 2005.

RIFAI N, GILLETTE MA, CARR SA: Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24**: 971-983, 2006.

RIGHETTI PG, BOSCHETTI E: Sherlock Holmes and the proteome--a detective story. *FEBS J* **274**: 897-905, 2007.

RIGHETTI PG, BOSCHETTI E, LOMAS L, CITTERIO A: Protein Equalizer Technology: the quest for a "democratic proteome". *Proteomics* **6**: 3980-3992, 2006.

ROSA DD, ISMAEL G, LAGO LD, AWADA A: Molecular-targeted therapies: lessons from years of clinical development. *Cancer Treat Rev* **34**: 61-80, 2008.

ROSS PL, HUANG YN, MARCHESE JN, WILLIAMSON B, PARKER K, HATTAN S, KHAINOVSKI N, PILLAI S, DEY S, DANIELS S, PURKAYASTHA S, JUHASZ P, MARTIN S, BARTLET-JONES M, HE F, JACOBSON A, PAPPIN DJ: Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154-1169, 2004.

ROUX-DALVAI F, GONZALEZ DE PEREDO A, SIMO C, GUERRIER L, BOUYSSIE D, ZANELLA A, CITTERIO A, BURLET-SCHILTZ O, BOSCHETTI E, RIGHETTI PG, MONSARRAT B: Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. *Mol Cell Proteomics* **7**: 2254-2269, 2008.

SCHIESS R, WOLLSCHEID B, AEBERSOLD R: Targeted proteomic strategy for clinical biomarker discovery. *Mol Oncol* **3**: 33-44, 2009.

SCHIRLE M, HEURTIER MA, KUSTER B: Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **2**: 1297-1305, 2003.

SCHMIDT A, KELLERMANN J, LOTTSPEICH F: A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**: 4-15, 2005.

SCHNOLZER M, JEDRZEJEWSKI P, LEHMANN WD: Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17**: 945-953, 1996.

SCHRODER FH, HUGOSSON J, ROOBOL MJ, TAMMELA TL, CIATTO S, NELEN V, KWIATKOWSKI M, LUJAN M, LILJA H, ZAPPA M, DENIS LJ, RECKER F, BERENGUER A, MAATTANEN L, BANGMA CH, AUS G, VILLERS A, REBILLARD X, VAN DER KWAST T, BLIJENBERG BG, MOSS SM, DE KONING HJ, AUVINEN A: Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**: 1320-1328, 2009.

SENNELS L, SALEK M, LOMAS L, BOSCHETTI E, RIGHETTI PG, RAPPSILBER J: Proteomic analysis of human blood serum using peptide library beads. *J Proteome Res* **6**: 4055-4062, 2007.

SHAH SJ, YU KH, SANGAR V, PARRY SI, BLAIR IA: Identification and quantification of preterm birth biomarkers in human cervicovaginal fluid by liquid chromatography/tandem mass spectrometry. *J Proteome Res* **8**: 2407-2417, 2009.

SHEVCHENKO A, JENSEN ON, PODTELEJNIKOV AV, SAGLIOCCO F, WILM M, VORM O, MORTENSEN P, SHEVCHENKO A, BOUCHERIE H, MANN M: Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci USA* **93**: 14440-14445, 1996.

SPIRO RG: Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R-56R, 2002.

STATES DJ, OMENN GS, BLACKWELL TW, FERMIN D, ENG J, SPEICHER DW, HANASH SM: Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **24**: 333-338, 2006.

STESSL M, NOE CR, LACHMANN B: Influence of image-analysis software on quantitation of two-dimensional gel electrophoresis data. *Electrophoresis* **30**: 325-328, 2009.

TAJIRI M, OHYAMA C, WADA Y: Oligosaccharide profiles of the prostate specific antigen in free and complexed forms from the prostate cancer patient serum and in seminal plasma: a glycopeptide approach. *Glycobiology* **18**: 2-8, 2008.

TAM SW, PIRRO J, HINERFELD D: Depletion and fractionation technologies in plasma proteomic analysis. *Expert Rev Proteomics* **1**: 411-420, 2004.

THIEDE B, HOHENWARTER W, KRAH A, MATTOW J, SCHMID M, SCHMIDT F, JUNGBLUT PR: Peptide mass fingerprinting. *Methods* **35**: 237-247, 2005.

THOMPSON A, SCHAFER J, KUHN K, KIENLE S, SCHWARZ J, SCHMIDT G, NEUMANN T, JOHNSTONE R, MOHAMMED AK, HAMON C: Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**: 1895-1904, 2003.

THULASIRAMAN V, LIN S, GHEORGHIU L, LATHROP J, LOMAS L, HAMMOND D, BOSCHETTI E: Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. *Electrophoresis* **26**: 3561-3571, 2005.

TIAN Y, ZHOU Y, ELLIOTT S, AEBERSOLD R, ZHANG H: Solid-phase extraction of N-linked glycopeptides. *Nat Protoc* **2**: 334-339, 2007.

TUCHOLSKA M, BOWDEN P, JACKS K, ZHU P, FURESZ S, DUMBROVSKY M, MARSHALL J: Human serum proteins fractionated by preparative partition chromatography prior to LC-ESI-MS/MS. *J Proteome Res* **8**: 1143-1155, 2009.

TUMANI H, PFEIFLE M, LEHMENSIEK V, RAU D, MOGEL H, LUDOLPH AC, BRETTSCHNEIDER J: Candidate biomarkers of chronic inflammatory demyelinating polyneuropathy (CIDP): Proteome analysis of cerebrospinal fluid. *J Neuroimmunol* **214**: 109-112, 2009.

UNLU M, MORGAN ME, MINDEN JS: Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071-2077, 1997.

UNWIN RD, GRIFFITHS JR, WHETTON AD: A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nat Protoc* **4**: 870-877, 2009.

VEENSTRA TD, CONRADS TP, HOOD BL, AVELLINO AM, ELLENBOGEN RG, MORRISON RS: Biomarkers: mining the biofluid proteome. *Mol Cell Proteomics* **4**: 409-418, 2005.

VILLANUEVA J, PHILIP J, ENTENBERG D, CHAPARRO CA, TANWAR MK, HOLLAND EC, TEMPST P: Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem* **76**: 1560-1570, 2004.

VILLANUEVA J, SHAFFER DR, PHILIP J, CHAPARRO CA, ERDJUMENT-BROMAGE H, OLSHEN AB, FLEISHER M, LILJA H, BROGI E, BOYD J, SANCHEZ-CARBAYO M, HOLLAND EC, CORDON-CARDO C, SCHER HI, TEMPST P: Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* **116**: 271-284, 2006.

WADA Y, AZADI P, COSTELLO CE, DELL A, DWEK RA, GEYER H, GEYER R, KAKEHI K, KARLSSON NG, KATO K, KAWASAKI N, KHOO KH, KIM S, KONDO A, LATTOVA E, MECHREF Y, MIYOSHI E, NAKAMURA K, NARIMATSU H, NOVOTNY MV, PACKER NH, PERREAULT H, PETER-KATALINIC J, POHLENTZ G, REINHOLD VN, RUDD PM, SUZUKI A, TANIGUCHI N: Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**: 411-422, 2007.

WASINGER VC, CORDWELL SJ, CERPA-POLJAK A, YAN JX, GOOLEY AA, WILKINS MR, DUNCAN MW, HARRIS R, WILLIAMS KL, HUMPHERY-SMITH I: Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **16**: 1090-1094, 1995.

WEI YS, ZHENG YH, LIANG WB, ZHANG JZ, YANG ZH, LV ML, JIA J, ZHANG L: Identification of serum biomarkers for nasopharyngeal carcinoma by proteomic analysis. *Cancer* **112**: 544-551, 2008.

WHITEAKER JR, ZHANG H, ZHAO L, WANG P, KELLY-SPRATT KS, IVEY RG, PIENING BD, FENG LC, KASARDA E, GURLEY KE, ENG JK, CHODOSH LA, KEMP CJ, MCINTOSH MW, PAULOVICH AG: Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J Proteome Res* **6**: 3962-3975, 2007.

WOLLSCHEID B, BAUSCH-FLUCK D, HENDERSON C, O'BRIEN R, BIBEL M, SCHIESS R, AEBERSOLD R, WATTS JD: Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**: 378-386, 2009.

WOLTERS DA, WASHBURN MP, YATES JR 3RD: An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**: 5683-5690, 2001.

WORKMAN P: The opportunities and challenges of personalized genome-based molecular therapies for cancer: targets, technologies, and molecular chaperones. *Cancer Chemother Pharmacol* **52** (Suppl 1): S45-S56, 2003.

XU G, XIANG CQ, LU Y, KANG XN, LIAO P, DING Q, ZHANG YF: Application of SELDI-TOF-MS to identify serum biomarkers for renal cell carcinoma. *Cancer Lett* **282**: 205-213, 2009.

YOUNG LS, SEARLE PF, ONION D, MAUTNER V: Viral gene therapy strategies: from basic science to clinical application. *J Pathol* **208**: 299-318, 2006.

YU KH, BARRY CG, AUSTIN D, BUSCH CM, SANGAR V, RUSTGI AK, BLAIR IA: Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery. *J Proteome Res* **8**: 1565-1576, 2009.

ZHANG H, LIU AY, LORIAUX P, WOLLSCHEID B, ZHOU Y, WATTS JD, AEBERSOLD R: Mass spectrometric detection of tissue proteins in plasma. *Mol Cell Proteomics* **6**: 64-71, 2007a.

ZHANG X, LI L, WEI D, YAP Y, CHEN F: Moving cancer diagnostics from bench to bedside. *Trends Biotechnol* **25**: 166-173, 2007b.

ZINKIN NT, GRALL F, BHASKAR K, OTU HH, SPENTZOS D, KALMOWITZ B, WELLS M, GUERRERO M, ASARA JM, LIBERMANN TA, AFDHAL NH: Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. *Clin Cancer Res* **14**: 470-477, 2008.