Biomarker discovery for cervical cancer
Govorukhina, Natalia I.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter I.II
Sample Preparation of Body Fluids for Proteomics Analysis

Natalia Govorukhina and Rainer Bischoff
This is an updated version of a book chapter entitled: Sample preparation of body fluids for proteomics analysis; in: Proteomics of human body fluids: Principles, Methods, and Applications, Ch. 2, 31-71 (2007); (Visith Thongboonkerd, Ed.), Humana Press (Totowa, New Jersey, USA).

1. Introduction
1.1. Proteomics of body fluids
1.2. Methodological overview

2. Sample preparation
2.1. The preparation of plasma and serum
2.2. Removal of high-abundance proteins
2.3. Enrichment of proteins or protein families
2.3.1. Lectins
2.3.2. Activity-based probes (ABPs)
2.4. Protein chip technology (SELDI-TOF-MS)
2.5. Magnetic beads and automated sample preparation
2.6. Methods adapted to other body fluids

3. The linkage to separation methods
3.1. 2D gel electrophoresis of proteins in body fluids
3.2. HPLC of proteins and peptides in body fluids
3.2.1. Peptidomics®
3.2.2. Shotgun proteomics

4. Synopsis
1. Introduction
1.1. Proteomics of Human Body Fluids

The analysis of human body fluids constitutes one of the most important approaches to the diagnosis of disease and in following therapeutic interventions. Human body fluids carry information about the status of the organism that may help in the recognition of physiological misbalances when overt pathological symptoms are not yet present. Analyzing the constituents of body fluids presents a number of challenges, the most difficult being the discrimination between variability in composition caused by an ongoing disease process and natural variability. The composition of body fluids varies due to endogenous, possibly pathological, processes and many environmental influences such as diet and lifestyle and the way the organism deals with them (e.g., metabolism and detoxification). This variability is most obvious when one is analyzing samples from different persons (cross-sectional studies) but is also present, albeit to a lesser extent, when one is analyzing samples from the same person over a given time period (longitudinal studies). Variability cannot be avoided but may be reduced by careful selection of the study population. At any rate, the discovery of disease-related changes in the composition of body fluids requires the study of a significant number of samples from patients and controls and a careful statistical interpretation of the results.

From an analytical chemistry point of view, body fluids constitute highly complex biological samples containing cells, proteins, peptides, and many metabolites. Thus, preparation of body fluids is unavoidable prior to determining the concentration or amount of a given set of constituents. Sample pretreatment and all further downstream steps will affect the ultimate analytical result and must therefore be carefully controlled and validated. It is not easy to give a general overview of sample pretreatments for body fluids, since each of them requires an adapted protocol, which in turn needs to be tailored to certain groups of analytes. In this chapter we focus on sample pretreatments for the analysis of proteins and peptides in serum. Although serum is just one example of a body fluid, albeit an important one, we will use it to highlight general principles of sample pretreatments that have a bearing on other kinds of body fluids.

The first step after taking a blood sample from a patient is to treat it in a way that makes it suitable for storage and subsequent analysis. A common initial step is to separate blood cells from soluble components, for example, by low-speed centrifugation. During sampling and centrifugation, it is pivotal to avoid disruption or activation of cells, notably hemolysis of red blood cells (which is shown by an orange to red color of the supernatant) and activation of platelets, which may lead to degranulation. The remaining supernatant, the blood plasma, may be stored as such in the case anticoagulants were added during collection to prevent blood clotting. Alternatively, blood clotting may be allowed or induced by leaving plasma at room temperature for a few hours.
Deciding whether to store plasma or serum for subsequent analyses is important. Although plasma is easier to prepare, it requires the presence of efficient anticoagulants for long-term stability. The components of the coagulation, fibrinolytic, and complement systems are all sensitive to contact with unnatural surfaces, such as plastic containers, glass, or injection needles, and there is a risk of activating these systems during processing steps (e.g., during chromatography or solid-phase extraction). The preparation of serum requires coagulation of the plasma, which is a complex biochemical process that may be difficult to control. In most hospital or laboratory settings, coagulation is effectuated at room temperature for 1 to 4 h. During this time the endogenous coagulation system is activated, leading to a cascade of proteolytic events that results in the formation of a fibrin-containing blood clot, which is usually removed by centrifugation.

It is obvious that activating a proteolytic system can have serious consequences for subsequent proteomic analyses, and some authors have noted that the coagulation time affects the resulting serum [1,3] However, proteolytic events associated with coagulation are highly controlled due to the sequence specificity of the major proteases (thrombin, factor Xa) and their well-defined location in the coagulation pathway (factor VIIIa, factor Xla) (Figure 1).

It is thus not clear whether the coagulation time affects the final composition of the proteome significantly, but there are indications that the lower molecular weight part, the so-called peptidome, is altered (Schulz-Knappe, personal communication).

In our initial studies, which applied tryptic digestion prior to LC-MS analysis (the shotgun approach), we have not observed major changes in the resulting profiles (Figure 2).

It is, however, important to validate this sample processing step carefully within the context of the overall analytical scheme (e.g., the complete protein vs the shotgun and peptidomics approach), because coagulation time is not well controlled in most laboratory or hospital settings and experience shows that it is hard to impose strict rules on hospital personnel with respect to this parameter. Finally, for retrospective studies on already acquired and stored serum samples, it is not possible to influence this step; thus the decision here is whether to include these samples in the analysis or not. Chapter IV of this thesis will deal in detail with the effect of clotting time on the serum protein profile.
Figure 1. Overview of the intrinsic and extrinsic coagulation pathways. Both pathways are activated during preparation of serum from blood plasma. Proteolytic activity associated with coagulation may affect the profile of low-molecular-weight proteins and peptides used for peptidomics analysis. F, factor; PK, protein kinase; PKa, protein kinase A; TAFI, tissue angiogenesis factor inhibitor; TFPI, tissue factor pathway inhibitor; TPA, tissue plasminogen activator; UPA, urokinase plasminogen activator. Reproduced with permission from Tapper H, Herwald H. Modulation of hemostatic mechanisms in bacterial infectious diseases. Blood 2000;96:2329–2337.
Figure 2. Correlation map of LC-MS data sets obtained from the analysis of serum (male) after depletion and tryptic digestion. Coagulation at room temperature was allowed to proceed for 1, 2, 4, 6, or 8 h. As indicated in this plot, high correlation of all data sets was found, with correlation coefficients above 0.8 in each case.

The way samples are initially treated determines in part what kind of analytes can be detected and quantified. Although this is true for any kind of analyte, it is particularly critical for the analysis of proteins and peptides, which are susceptible to degradation, precipitation, chemical modification (e.g. oxidation), adsorption to the walls of containers, and so on. Establishing a well-controlled and reproducible sampling procedure is therefore critical for any study involving human body fluids [4-7]. The proteomes in body fluids differ significantly from intracellular or tissue-derived proteomes, which are the subject of most proteomics studies. Systemic body fluids, like blood, sample the whole organism and give an average picture of the physiological state of that organism at a given point in time. Notably, blood contains a few high-abundance proteins that are to a large extent produced and secreted by the liver. In contrast, urine is a much more dilute body fluid that samples the
metabolic end products from blood. Its composition is greatly influenced by the status of the kidneys. Although every body fluid presents particular challenges with respect to sample pretreatment, it is fair to say that blood is one of the most difficult body fluids to analyze.

In the following, we will highlight a number of options for sample pretreatment prior to proteomics analysis. Our focus will be directed at serum, but the principles are applicable to other body fluids. We will try to emphasize that there are strategic choices to be made early on in the analytical procedure that will determine the final result.

1.2. Methodological Overview

There is no single approach to proteomics in body fluids. It is likely that the comprehensive analysis of the proteome of any given body fluid is still beyond our reach despite great methodological advances in recent years. A major challenge is the concentration range of proteins in most body fluids, which spans more than 11 orders of magnitude [8] (Figure 3).

![Figure 3. Concentration range of proteins found in human plasma. It is noteworthy that there is a difference of more than 11 orders of magnitude between the most concentrated and the very low-abundance proteins. Reproduced with permission from Anderson NL, Anderson NG. The Human Plasma Proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1:845–867.](image)
Furthermore, it is difficult to predict the number of proteins in body fluids owing to processing events (like the generation of smaller proteins and peptides from larger precursors, posttranslational modifications, and the fact that proteins can enter body fluids by well-defined pathways like secretion) as well as cell and tissue turnover as a result of necrosis or apoptosis. From a methodological point of view, the proteome of a body fluid may be roughly divided into high- and low-molecular-weight compartments (Figure 4).

Figure 4. Schematic overview of different approaches of sample preparation. The main initial difference is whether undigested or trypsin-digested proteins are being analyzed. A second level of differentiation comes into play when one is deciding whether to analyze the high- or the low-molecular-weight (MW) fraction of the proteome. High-molecular-weight proteins are best analyzed by 2D electrophoresis, whereas peptides are more amenable to high-performance liquid chromatography (HPLC) coupled to mass spectrometry.

This discrimination is rather arbitrary and is mostly defined by the size-dependent separation method used for prefractionation.

Body fluids have been prefractionated by ultrafiltration with membranes of various cutoff values [9-11]. Although ultrafiltration appears to be an easy separation methodology with an apparently clear-cut separation mechanism, its application to complex biological samples shows that discrimination between proteins above and below the nominative cutoff of the membrane is never complete [11]. Effects such as adsorption of proteins to the membrane, the generation of a polarization layer close to the membrane surface, and deformation of the pores in relation to the g-force can all affect filtration.
Ultrafiltration is also performed on a large scale in patients with renal insufficiency, and membrane material has been the source for many studies of bioactive proteins and peptides below approximately 20 kDa [12-14]. Although the kidney itself is performing much more complex tasks than just ultrafiltration, urine may be considered an ultrafiltrate of blood and thus has a considerably lower concentration of high molecular-weight proteins.

An elegant combination of ultrafiltration and chromatography is based on restricted access materials (RAMs), which have an adsorbing internal pore surface and a non-adsorbing external surface [15,16]. The pore diameter in most RAMs is about 6 nm, which corresponds roughly to a cutoff value of 20 kDa. RAM chromatography has been integrated into analytical systems with the goal of analyzing the low-molecular-weight part of the proteome of blood diafiltrates [17,18] or artificial cerebrospinal fluid [19]. Although a clear enrichment of the low-molecular-weight fraction was observed, a considerable amount of albumin was still present even after RAM chromatography.

The decision whether to work with undigested proteins or to digest proteins with trypsin prior to further analyses is of principal importance in an analytical scheme (Figure 4). Performing separations of very complex mixtures of proteins is difficult, owing to the wide range of physicochemical properties and the possibility that proteins will denature, aggregate, or even precipitate under separation conditions. The most universally applicable separation method for proteins is 2D polyacrylamide gel electrophoresis (2D-PAGE), whereby all proteins are denatured from the beginning and kept in a denatured state throughout separation. This reduces the risk of aggregation and precipitation with subsequent loss of proteins as well as that of proteolysis. There is no comparable universal chromatographic method, and it is thus necessary to develop an appropriate fractionation scheme for groups of proteins or individual proteins. However, as 2D-PAGE has limitations with respect to low-molecular-weight proteins of 10 to 20 kDa and hydrophobic or basic proteins, alternatives are being developed.

One approach is based on the so-called shotgun method, whereby the complete protein mixture is digested with trypsin (other proteases are conceivable as well for this purpose but are not widely used) and the generated peptides are separated by 2D or 3D high-performance liquid chromatography (HPLC) [20-23]. Shotgun proteomics has the advantage of overcoming many of the difficulties related to very hydrophobic or otherwise intractable proteins at the expense of rendering the separation problem quite daunting. Assuming that serum contains about $10^5$ different protein forms, each of which generates 50 tryptic peptides, one has to deal with a mixture of about $5 \times 10^6$ peptides to be separated. Fortunately, not all peptides need to be separated into single peaks, and not all peptides of each protein need to be identified by mass spectrometry to trace them back to the protein of origin. A drawback of the shotgun method is that not all regions of a protein are covered by the analysis,
which may mean that some possibly relevant posttranslational modifications or processed forms are missed. Nevertheless, the excellent separation capacity of HPLC for peptides compared with complete proteins and the much easier identification of peptides by tandem mass spectrometry have accelerated the use of shotgun proteomics in the biomarker discovery area. The daunting separation problem posed by this approach has also driven recent new developments in HPLC stationary phase chemistry and technology that increase separation efficiency and reduce analysis time [24-28].

The presence of a few high-abundance proteins in body fluids such as albumin has driven developments to deplete these proteins specifically and thus to increase the loading capacity by a factor of 5 to 10 [29-35] or even 30 to 50, depending on the number of proteins that are depleted (e.g. ProteoPrep® 20 Plasma Immunodepletion kit, www.Sigma-Aldrich.com removes the 20 most abundant serum proteins). In addition to increasing the protein or peptide load, depletion also augments the capacity to detect peptides derived from lower abundance proteins [36]. It has been shown, however, that depletion of high-abundance proteins does not proceed without the loss of some low-molecular-weight proteins and peptides [37]. It is thus important to decide whether to deplete or rather try to design a fractionation strategy that deals with high-abundance proteins such as albumin or immunoglobulins in another way [38].

Arguably, very low-abundance proteins in the ng/mL to pg/mL range cannot be detected in complex protein mixtures such as serum even after depletion. Many regulatory proteins such as cytokines or some of the known tumor-specific markers reside in this concentration range and are presently measured by immunological methods. To reach into this lower concentration range, it is often necessary to enrich a given set of proteins by affinity chromatography using highly selective antibodies or group-specific ligands like lectins. The use of protein-specific antibodies limits the scope of the analysis to those proteins that are recognized. Group-specific affinity ligands such as lectins or antibodies directed at a common structural element such as phosphotyrosine represent a compromise between the comprehensive proteomics approach, which often fails to detect low-abundance proteins, and the highly specific methods. For example, lectins have been applied to the discovery of tumor-specific glycoprotein markers, since tumor cells often produce proteins with aberrant glycosylation patterns [39-41]. Lectins have also been used to enrich glycoproteins from complex protein mixtures or glycopeptides from tryptic digests of such mixtures [42-45]. In general, targeted approaches require a hypothesis concerning the role that different kinds of proteins may play in a given disease in order to chose appropriate affinity ligands for enrichment.
2. Sample Preparation
2.1. Preparation of Plasma and Serum

Between two fundamentally different compartments of the blood, namely, blood cells and the actual fluid, most clinical analyses are done on derivatives of the fluid, like plasma or serum. Discussion continues on whether serum or plasma should be used, but this may also depend on the general practice of the hospital that provides samples for analysis, notably, whether samples are analyzed from existing collections. Preparation of plasma requires addition of anticoagulants, such as EDTA, citrate, and/or heparin, whereas serum contains no extra additives. Serum lacks components of the coagulation system, such as thrombin and fibrinogen, since they are part of or become entrapped in the blood clot and are thus removed from the serum. In addition, other proteins or peptides that have some binding affinity to the clot may be partially depleted. Being a proteolytic process, coagulation generates peptide fragments from larger proteins that may especially affect the composition of the peptidome. Interestingly, comparison of plasma made with EDTA, citrate, or heparin also shows variation in protein composition [46].

Sampling blood for plasma or serum preparation is routine in most hospital laboratories and a reasonably standardized procedure is in place using commercial reagents and materials. However, most laboratory technicians and nurses are not aware of the specific requirements of proteomics and thus need to be informed. Very restrictive standard operating procedures (SOPs), notably with respect to the coagulation time and conditions, are often difficult to follow in routine hospital operations. Biomarkers discovered thus far therefore need to be robust enough to be useful in a routine clinical laboratory, and very unstable proteins or peptides are probably not of interest in the long run.

Recently, Schulte et al. [2] reported that a considerable number of peptides were found in serum but not in human plasma. The authors suggest that these peptides appeared as a result of a clotting-related proteolytic activity. This might be indicative of artifacts generated as a result of the clotting reaction, which is disturbing with respect to peptidomics. The authors therefore propose to use human plasma for this purpose.

An example of the preparation of plasma for biomarker discovery by Peptidomics® (Schulz-Knappe, personal communication) involves taking a blood sample from a superficial vein of the cubital region. The blood sampling procedure should not take longer than 1 min, and EDTA is used as the anticoagulant. Prior to collection, the first sample (approximately 2.5 mL) is discarded. To remove platelets, the sample is centrifuged at 2,000 g for 10 min. The final plasma sample (approximately 1.5 mL) should be frozen within 30 min after being taken and stored at -80°C. Serum is made by letting a fresh blood sample coagulate (with or without thrombin as activator) and either filtering it through a gel or collecting the liquid fraction after centrifugation. Although their hypothesis has not been proven, Sorace and Zhan [47] suggest
that variations in coagulation might be a significant factor in obscuring clinical proteomics data sets. The source of variation can be both technical and natural. Schulte et al. [2] found that a naturally occurring Val-34 to Leu mutation in the activation peptide of factor XIII (FXIIIA) not only affected the process of blood clotting but also correlated with a lower incidence of myocardial infarction and ischemic stroke and an increased risk of hemorrhagic stroke. According to our results (see chapter IV of this thesis), different clotting times ranging from 1 to 8 h in the preparation of serum samples resulted in highly correlated liquid chromatography-mass spectrometry (LC-MS) data sets when analyzing serum proteins after depletion and trypsin digestion. Correlation coefficients above 0.8 were found for all samples after selecting the 37 top information-rich m/z traces using the CODA component detection algorithm [48] (Figure 2).

2.2. Removal of High-Abundance Proteins

As the presence of abundant proteins in most biofluids used for diagnostic purposes decreases the capacity of analytical methods to detect low-abundance proteins or peptides, a range of approaches has been developed to reduce the total amount of protein. Blood serum is a complex mixture of thousands of proteins and peptides. However, few of the serum proteins are present in extremely high amounts compared with the rest of the serum components. (human serum albumin [HSA] constitutes 57–71% and γ-globulins 8–26% of the total of all human serum proteins). The 10 most abundant proteins account for 97% of all the protein content in plasma [8]. In a recent publication, it is stated that the search for specific markers occurs in a fraction of less than 1% of all plasma proteins [49].

Removal of these high-abundance proteins increases the loading capacity of the analytical system by a factor of 5 to 10 and thus improves the detection of low-abundance proteins. Several affinity columns are presently on the market based on dye ligands or antibodies for albumin removal and protein A or G for the removal of immunoglobulins [50]. Technically simple approaches that allow processing of multiple samples in parallel based on HSA- and IgG-binding spin columns or filters have been developed [51]. For HSA binding, two types of stationary phases are generally used: (1) those based on dye ligands such as Cibacron-Blue and derivatives thereof [52], and (2) those based on specific antibodies against human serum albumin [51] raised in mammals (IgG) but also in chickens (IgY), as recently described [31]. HSA was also successfully removed by affinity capture on immobilized phage-derived peptides [53]. Recently, a synthetic peptide derived from protein G was used for HSA affinity chromatography and depletion of HSA from human plasma [54]. The column could easily be regenerated with alkaline treatment owing to the stability of the peptide, and its specificity and capacity were quite high. However, the column is presently not commercially available. Removal of IgG is exclusively done by well-established methods based on immobilized protein
A, protein G, or protein L, owing to their high affinity and selectivity [55-60]. Comparative studies of HSA- and IgG-binding columns based on Poros® polystyrene-divinylbenzene beads (Applied Biosystems) [32] as well as on Mimetic Blue (ProMetic BioSciences) and HiTrap Blue (Amersham Biosciences) for HSA removal have been performed [33].

We tested several approaches specifically to reduce the level of high-abundance proteins in serum based on either specific antibodies, dye ligands (for albumin), or protein A or G (for γ-globulins) [30]. Analysis by sodium dodecyl sulfate (SDS)-PAGE (Figure 5) and LC-MS after tryptic digestion of the remaining proteins (Figure 6), showed that reduction with albumin-directed antibodies was most effective, albeit not complete [36].

![Image of SDS-PAGE gels](image.png)

Figure 5. Depletion of albumin and γ-globulins from human serum. In each lane 8 to 10 μg of protein were loaded, and gels were stained with Coomassie Blue G-250. (A) POROS Anti-HSA and POROS Protein G columns. (B) HiTrap Blue and HiTrap Protein G columns. (C) Merck Albumin Removal column and HiTrap Protein G column. (D) Aurum Serum Protein column. Lanes: 1, standards; 2, crude serum; 3, depleted serum; 4, bound protein eluted from albumin-depleting columns; and 5, bound γ-globulins eluted from columns. Labeled protein bands: 1, serotransferrin; 2, α1-antitrypsin; 3, albumin; 4, 5, γ-globulins, heavy and light chains, respectively. Reproduced with permission from Govorukhina NI, Keizer-Gunnink A, van der Zee AGJ, de Jong S, de Bruijn HWA, Bischoff R. Sample preparation of human serum for the analysis of tumor markers: comparison of different approaches for albumin and [gamma]-globulin depletion. J Chromatogr A 2003;1009:171-178.

In our initial studies, which applied tryptic digestion prior to LC-MS analysis (the shotgun approach), we have not observed major changes in the resulting profiles (Figure 2).

A more recently introduced multiple affinity removal column, which depletes certain high-abundance proteins (albumin, IgG, IgA, transferrin,
haptoglobin, and α₁-antitrypsin [50]), proved to be most effective in our hands and provided more reproducible results during LC-MS analysis regarding retention times and peak areas than previously evaluated methods [61,62]. Similar results were recently published for 2D gel electrophoresis (Figure 7) [50]. Recently immunoaffinity materials have been introduced that remove the 20 most abundant proteins (ProteoPrep® 20 Plasma Immunodepletion kit/ spin columns (www.Sigma-Aldrich.com)) or the 14 most abundant proteins (Hu-14 Spin Cartridges (www.home.agilent.com)).

Figure 6. Efficiency (A) and selectivity (B) of albumin removal from human serum using an anti-albumin immunoaffinity column. (A) Extracted ion chromatogram of $m/z = 575.3$ (doubly charged molecular ion of peptide LVNEVTEFAK; positions 41–50 in human serum albumin) of tryptic digests of human serum (upper trace; peak height $3.2 \times 10^7$) or of human serum after depletion with an anti-albumin immunoaffinity column (lower trace; peak height $4.6 \times 10^5$). (B) Extracted ion chromatogram of $m/z = 393.3$ (doubly charged molecular ion of peptide IVDLVK; positions 193–198 in human α₁-antitrypsin) of tryptic digests of human serum (upper trace; peak height 16,052) or of human serum after depletion with an antialbumin immunoaffinity column (lower trace; peak height 32,607). Note the much cleaner detection of this peptide fragment after depletion and the increased overall peak height. Reproduced with permission from Bischoff R, Luider TM. Methodological advances in the discovery of protein and peptide disease markers. J Chromatogr B 2004;803:27–40.
Figure 7. 2D electrophoresis of crude (70 μg protein) and depleted (6-protein multiple affinity removal column, 100 μg protein) serum samples. On the right side is a zoom view of the area containing albumin. Reproduced with permission from Bjorhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* 2005;5:307–317.

Removal of high-abundance proteins by ultrafiltration through cellulose filters with a cutoff of 30 kDa proved to be less successful [11]. Many known “landmark” proteins of low molecular weight (<30 kDa) were missing upon 2D gel electrophoresis. Ultrafiltration has, however, the advantage of allowing one to concentrate the low-molecular-weight fraction of the proteome and was found to be useful for analysis of low-molecular-weight proteins (LMWs) by LC-MS after prefractionation by strong cation-exchange HPLC [10]. To prevent binding of LMWs to serum carriers, particularly albumin, 20% acetonitrile was used. In the resulting fraction, the authors could identify 314 unique proteins including cytokines, growth factors, and transcription factors, which are proteins of low abundance that are very difficult to detect by other methods without serum depletion. This method of sample preparation at a cutoff of 10 kDa was recently used to analyze the LMW fraction of pooled serum from patients with ovarian cancer by nanoLC-electrospray ionization-fourier
A quite different set of methods uses electrophoretic approaches to fractionate complex samples and to separate high-abundance proteins from those of low abundance. The basic principle is based on preparative isoelectric focusing and/or free-flow electrophoresis in solution, whereby the crude sample is prefractionated in a specially designed chamber according to the different electrophoretic mobilities or isoelectric points of the proteins. The HSA-rich fraction was discarded, and other fractions were pooled or analyzed separately. The method was originally reported more than 10 years ago [63] and is still in limited use [64-66]. Some commercially available systems can be used for electrophoretic prefractionation (e.g., the Zoom IEF fractionator® [Invitrogen] or the system produced by Weber).

There is one particular problem associated with the removal of serum albumin and globulins. These proteins appear to fulfill the function of carriers for less abundant proteins [37,67]. This is especially critical for LMWs, since they can escape kidney clearance only when bound to high-molecular-weight carrier proteins. Many of these LMWs are found to be associated with the development of cancer and could therefore be extremely important biomarkers (see ref. 37). Binding of LMWs to high-abundance, high-molecular-weight proteins may be used advantageously based on a two-step procedure, whereby abundant carrier proteins are first specifically bound to the corresponding affinity column followed by elution of the bound LMWs using a gradient [37,67].

2.3. Targeted Enrichment of Individual Proteins or Protein Families

Since many disease-specific biomarkers are likely of low to very low abundance in body fluids, it is a major challenge to detect them using profiling methods. Reaching the required concentration sensitivity often requires complex, well-designed protocols of enrichment and fractionation that are rather time-consuming. A concept of a multi-dimensional fractionation system (MDFS) was recently proposed and discussed [68]. Based on a given hypothesis about the disease mechanism, it is therefore often advisable to use targeted, affinity-based methods for enrichment prior to analysis. A combination of proteomics technology with targeted enrichment that does not require a very “sharp” hypothesis is based on group-specific ligands like lectins [69-70] in case of glycoproteins, or activity-based probes (ABPs) in the case of proteases or other enzymes [71-77].

2.3.1. Lectins

Glycosylation of proteins is a posttranslational modification that is easily affected by cellular growth conditions. Modifications of glycosylation patterns are therefore often observed in fast growing cancerous cells compared with
their quiescent counterparts [39,78]. Analysis of the carbohydrate portion of proteins is a rather complex task, since the glycosyl moiety is usually a branched chain polymer with an enormous variety in length, composition, and complexity. Studies of glycoproteins can be divided into two types: 1) identification of the proteins and their glycosylation sites and 2) the more demanding analysis of the structure of the attached glycosyl residues themselves.

Glycosyl residues can be linked to the protein core via asparagine (N-linked glycans) or bound via serine or threonine (O-linked glycans). For N-linked glycans, N-acetylglucosamine (GlcNAc) is the first monosaccharide in the chain, whereas N-acetylgalactosamine (GalNAc) is most often found for O-linked glycans. In addition to being potentially interesting as biomarkers, failure of proper glycosylation can cause severe abnormalities [79].

Although the exact structure of glycosylated proteins varies considerably, probably all known glycoproteins can be enriched by lectin affinity chromatography [80]. There are several commercially available lectin affinity columns, which differ in specificity and are used widely in early stages of the isolation of glycoproteins [81]. The specificity of many lectins is known (Table 1), allowing the rational design of complementary enrichment schemes.

**Table 1: Selected lectins with their specificities.**

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concanavalin A (ConA)</td>
<td>glucosyl and mannosyl residues of N-linked oligosaccharides</td>
</tr>
<tr>
<td>Wheat germ agglutinin (WGA)</td>
<td>chitobiose core (di-N-acetylglucosamine) and N-acetylneuraminic acid</td>
</tr>
<tr>
<td>Peanut agglutinin (PNA)</td>
<td>T antigen (Galb1-3GalNAc) found in O-glycans of mucin-type proteins</td>
</tr>
<tr>
<td>Aleuria aurantia (AAL)</td>
<td>L-fucose–containing oligosaccharides</td>
</tr>
<tr>
<td>Galectines</td>
<td>N-acetyllactoseamine (LacNAc)-containing glycans found in both N- and O-glycans</td>
</tr>
</tbody>
</table>

Recently five lectins, concanavalin A (Con A), wheat germ agglutinin (WGA), jacalin, lentil lectin (LCA), and peanut lectin (PNA) were tested for capturing glycoproteins from human serum [41]. At first, the authors depleted the high-abundance proteins with a multiple affinity depletion column followed by enrichment on single- or multiple-lectin affinity columns. The enriched proteins were eluted with buffers containing specially selected sugars. The resulting fractions were analyzed by LC-MS after digestion with trypsin. Figure 8 gives an example of how the analysis of proteins in serum can be focused to a subset containing a fucose residue by prior enrichment on a column containing the fucose-specific Lotus tetragonolobus agglutinin (LTA) [43].
2.3.2. Activity-Based Profiling of Proteases

Standard proteomics techniques give information about the relative abundance of proteins and possibly posttranslational modifications. In most cases, however, these techniques do not provide information about biological activity. In recent years another branch of proteome analysis has developed to tackle this problem with the development of affinity-labeling techniques, generally called activity-based protein profiling (ABPP) [82-85]. This line of research focuses on profiling the activity of families of enzymes like the various types of proteases. A derivative of this work is to use affinity ligands, like protease inhibitors, to enrich classes of proteins based on their activity [76]. Arguably, it is the activity of enzymes that is involved in disease development and that may therefore serve as biomarkers rather than the abundance, since most enzymes are present as inactive proforms that are activated upon appropriate (or inappropriate) stimuli.

ABPs have been described for cysteine proteases [84,86-88], serine hydrolases, including serine proteases [73,89], and also metalloproteases [71,90-92]. In most cases the labels contain biotin, which allows one not only to visualize but also to isolate the labeled proteins. Even in vivo labeling, for example, in tissue biopsies or cells in culture is feasible [84]. By employing this strategy, sample depletion for abundant proteins can be bypassed as long as the inhibitors or other affinity ligands are sufficiently specific and nonspecific.

Figure 8. Enrichment of fucose-containing peptides in serum proteins after tryptic digestion. An LTA lectin affinity column was used to select the peptides followed by deglycosylation and reversed-phase HPLC. Reproduced with permission from Xiong L, Andrews D, Regnier F. 2003. Comparative proteomics of glycoproteins based on lectin selection and isotope coding. J Proteome Res 2003;2:618–625.
binding to the support materials can be minimized. Figure 9 gives an example of how strongly some proteins may bind to materials used for the immobilization of affinity ligands and how nonspecific binding may be overcome by chemical derivatization of the surface.

![Figure 9](image)

Figure 9. Binding of the catalytic domain of matrix metalloprotease-12 (MMP-12) to unmodified controlled porosity glass beads (▲), silanized with a diol layer (●) or silanized and then coupled to a reversible MMP inhibitor (■) (Dr. Robert Freije, unpublished data).

Nonspecific binding to affinity ligands or the surfaces on which they have been immobilized makes stringent controls necessary. For example, preparing non-functionalized “control” materials and competing with the affinity interaction by adding an excess of ligand to the binding buffer are common ways of assessing specificity. Some authors have also denatured the proteins by heat treatment prior to binding as a control. As an example, a complex mixture of extracellular serine hydrolases was successfully identified by MS after the hydrolases were enriched by affinity chromatography. In addition, resolution of protein bands on SDS-PAGE was improved upon deglycosylation of the enriched enzymes [93,94].

Successful affinity-based profiling requires suitable affinity ligands. To address a wider range of proteins based on their activity, combinatorial chemistry approaches have been developed [95]. A new promising concept of in vivo click chemistry protein labeling utilizing the copper(I)-catalyzed azide-alkyne cycloaddition reaction was recently introduced [96]. However, this method is mostly limited to catalytically active proteins, such as enzymes. A great number of protein-protein interactions, however, remain largely unexplored.
2.4. Protein Chip Technology (SELDI-TOF-MS)

Since natural body fluids are too complex to analyze directly, investigators are in constant search of new techniques of subfractionation prior to MS. Most often prefractionation is done by LC or electrophoresis, but simple adsorption/washing/desorption methods are finding more widespread use, as they are rather fast and can be automated more easily. In classical matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), prefractionated protein samples are digested with trypsin prior to analysis (peptide mass fingerprinting). In this version all peptides are indiscriminately deposited on the MALDI target plate and entrapped in the light-absorbing matrix.

The central idea of surface-enhanced laser desorption/ionization (SELDI) is to use adsorptive surfaces, mainly based on well-known chromatographic principles, to bind a subfraction of proteins from a sample and to analyze the bound proteins or peptides by MALDI-MS. By varying the adsorptive surface, different groups of proteins can be bound and analyzed. This technology has more recently been further developed and commercialized under the trade name ProteinChip® (Ciphergen Biosystems, Palo Alto, CA) and has found widespread application, notably in the medical and clinical research community [97]. The original mass spectrometer was a simple linear MALDI-time of flight (TOF) system, but interfaces have now been developed that allow coupling of the ProteinChip technology to tandem mass spectrometers of the quadrupole-TOF hybrid type.

The key components of this technology are Protein Chip Arrays and the Protein Chip Reader. The array comprises a set of different surfaces, such as a hydrophobic, hydrophilic, or metal chelate to which the biological samples are added. The unbound proteins are washed away, and the bound fraction is subjected to MS analysis. Optionally, it is possible to digest the bound proteins with trypsin to facilitate their identification. However, since most chips bind many diverse proteins, interpretation of the results after trypsin digestion is not always obvious because it is not straightforward to link the observed peptides back to the proteins that gave rise to an increased or decreased peak in the original spectrum. As a recent example of this approach, cystatin C was suggested as a biomarker in the diagnosis of Creutzfeld-Jakob disease [98]. Direct fragmentation of the high-molecular-weight ions to obtain sequence information for identification would be most advantageous. This has recently been facilitated due to ion activation mechanisms such as Electron Capture/Transfer Dissociation or Infrared Multiphoton Dissociation [99-104].

As with any mass spectrometric method dealing with highly complex mixtures, there is a competition between different molecules to ionize (also known as ion suppression). It is thus unlikely that the mass spectrum obtained from a ProteinChip will give a true representation of the proteins or peptides adsorbed on the chip. Most applications of SELDI-TOF-MS to body fluids
therefore generate rather simple mass spectra, which can be easily analyzed. A number of applications of SELDI-TOF to biological samples, notably plasma, serum, or urine, have shown that samples taken from patients differ significantly from those from healthy controls or from patients with other kinds of disease, opening the possibility of discriminating patient groups and performing early diagnosis of, for example, ovarian or breast cancer [105-106]. However, recent efforts to reproduce these results have met with limited success, and the jury is still out on whether this fairly straightforward approach to sample preparation will lead to clinically relevant results [107-109].

Probably one of the most impressive studies using SELDI-TOF in recent years was the detection of an antiviral factor secreted by CD8 T-cells upon infection with HIV-1 in immunologically stable patients, which was identified as a member of the α-defensin family by subsequent isolation and protein sequencing [110]. This factor had been known since 1986 but had eluded identification for 15 years [111].

2.5. Automated Sample Preparation Using Magnetic Beads

The automation of sample preparation in light of increasing sample throughput and reproducibility is an important aspect of clinical proteomics. In analogy to the previously described ProteinChips, it is possible to prepare samples by adsorption/washing/desorption on magnetic beads (or other kinds of beads). Magnetic beads are an effective tool for fast concentration of diluted samples and for the crude separation of proteins and peptides prior to MS analysis. Magnetic beads are widely used in automated immunoassays, cell purification, and more recently, the detection of bacterial pathogens. Mostly, the assay is targeted at individual proteins, like prostate-specific antigen (PSA), which is captured with a biotinylated anti-PSA antibody (anti-F-PSA-M30-IgG) and subsequently bound to streptavidin-coated magnetic beads [112]. This approach, however, is targeted to a specific protein and is not applicable to proteomic studies in a broad context. In another example, magnetic nanoparticles modified with vancomycin were used to trap Gram-positive bacteria [113]. The method was able to detect *Staphylococcus aureus* in a 3-mL urine sample at a concentration of $7 \times 10^4$ CFU/mL (colony-forming units) by MALDI-TOF-MS. A combination of affinity trapping with MS has also been successful in detecting bacterial and viral infections based on immobilized lectins [114,115]. Application of magnetic beads to clinical proteomics has emerged only recently mainly based on adapted liquid handling systems [116]. Serum was precipitated with ethanol to remove larger proteins, and the remaining polypeptides in the supernatant were bound to reversed-phase super-paramagnetic silica beads. The washed and eluted peptides were profiled by MALDI-TOF/TOF-MS with the possibility of performing partial sequencing and identification by MS/MS. Four hundred polypeptides were detected in 50 μL serum (range 0.8–15 kDa), and discrimination between
samples from brain tumor patients and healthy controls was 96.4% based on a learning algorithm. The number of examples of protein enrichment with magnetic beads has been steadily increasing over the last years [117-120].

2.6. Analysis of Other Body Fluids

Sample preparation is equally important for proteomics and peptidomics in other body fluids, and many of the methods and considerations developed for serum are suitable. Urine is probably the second most relevant body fluid after blood for general proteome studies, owing to its availability. Urine is in fact filtered blood plasma, so it might be representative of the protein spectrum of blood, but with lower protein concentrations. In normal conditions the kidney restricts passage of plasma proteins above approx 40 kDa during filtration in the glomeruli. Proximal renal tubules reabsorb filtered proteins and degrade them. Total amounts of secreted protein per voiding vary from 1 to 10 mg, whereas in pathology, the protein concentrations can dramatically increase [121,122]. Urine collects the metabolic end products of the organism destined for excretion, and its composition is therefore more variable than that of serum or plasma. In particular, the composition and concentration of proteins and peptides in urine are strongly affected by nutrition, the day/night cycle, and the health status of the kidneys. It is thus important to try to control and document these parameters as accurately as possible.

Although protein concentration in urine is much lower than in serum (approximately 1000-fold) and filtration takes place in the kidneys, albumin is still the major protein. Proteome maps of human urinary proteins were recently constructed after LC-MS analysis of trypsin-digested unfractionated urine [123,124], by 2D electrophoresis after acetone precipitation [125], and after depletion of high-abundance proteins (albumin and IgG) by ultrafiltration and 2D electrophoresis [126]. An equivalent of urine, human hemofiltrate, was also analyzed by restricted access chromatography to select the peptidome followed by 2D HPLC and MS [17,18]. Urine has furthermore been analyzed by capillary electrophoresis coupled to electrospray ionization (CE-MS) [127,128]. Combined with new analysis software, this analytical method is presently under further investigation. Normalization of the data obtained to an internal standard that takes biological variation into account is critical for urine [128]. This has been common practice in clinical chemistry for a long time, and creatinine is widely used for this purpose. However, whether creatinine is also a suitable normalization standard for proteomics and peptidomics studies in urine is questionable. Normalization based on the total protein content or the area under the curve of the HPLC-UV trace may be preferable. It remains to be seen whether the urinary spectrum of proteins and peptides can be successfully used to detect human diseases short of those related to the kidney or general inflammation.
Urine samples should be collected under sterile conditions, cooled down, and treated with protease inhibitors. The next steps of sample pretreatment vary from one publication to the next. For example, proteins can be concentrated by precipitation with trifluoroacetic acid followed by centrifugation [129]. The resulting sample can be further applied to 1D or 2D electrophoresis, or subjected to solid-phase extraction and trypsin digestion followed by LC-MS analysis. Pieper et al. [126] compared urinary proteomes of healthy and renal cell carcinoma patients. Initially, cooled samples with added protease inhibitors were cleared by centrifugation and concentrated by membrane filtration. Samples were further desalted and fractionated by gel filtration on Superdex G-75. The resulting sample of more than 30 kDa proteins was passed through a depleting column specific for albumin, IgG, and α-1-acid glycoprotein. The final comparative analysis was done by 2D electrophoresis and mass spectrometry. Urine samples were recently used for comparative studies of normal and lung cancer patients [130]. The collected urine samples were first desalted by gel filtration (PD-10 columns) followed by lyophilization. The pellet was resuspended in phosphate buffer, extracted with methanol/chloroform, and precipitated with trichloroacetic acid/acetone to remove organic acids and lipids. Finally, the sample was fractionated with HPLC and 1D or 2D electrophoresis followed by MALDI-MS and MS-MS. The image analysis of the gels demonstrated a quite impressive number of protein spots, but albumin and IgG were still quite abundant. Whereas easily obtainable body fluids such as blood, urine, saliva, or tears are samples of first choice for human proteomic studies, more specialized samples are frequently used to evaluate the condition of a given organ system. It is implied that a sample taken closer to the diseased organ will show changes in protein composition that are more closely related to the disease than those occurring in blood or urine.

Recently, the proteomics of bronchoalveolar lavage fluid (BALF) was reviewed [131,132]. Bronchoalveolar lavage samples the epithelial lining of the lung and is frequently analyzed in cases of severe respiratory diseases (e.g. chronic obstructive pulmonary disease, severe asthma, pulmonary fibrosis). BALF consists of a soluble part often used for biomarker analysis and cells derived from lung tissue or blood (alveolar macrophages, lymphocytes, neutrophils, and eosinophils). It is noteworthy that most proteins found in BALF correspond to abundant plasma proteins, indicating “plasma leakage” into the alveolar space owing to the lavage procedure. A map of the BALF proteome showed up to 1400 different proteins on a 2D gel [131], with some proteins at higher concentrations than in serum or plasma. These proteins are likely directly derived from the lung. Removing albumin as the most abundant protein in BALF by RAM chromatography allows one to process larger volumes and thus to detect lower abundance components (Figure 10).
Figure 10. Sample preparation of bronchoalveolar lavage fluid (BALF) by restricted access material (RAM) chromatography. The upper panel shows the reversed-phase HPLC analysis of 10 μL BALF (major peak is albumin), and the lower panel shows the analysis of 1 mL BALF after sample preparation (Dr. Begona Barroso, unpublished data).
3. The Linkage to Separation Methods and Mass Spectrometry

The analysis of complex proteomes requires that dedicated and effective sample preparation be followed by high-resolution separation to reduce complexity to a level that can be handled by MS in terms of protein or peptide ionization, identification, and quantification. A wide range of separation methods has been applied to proteins and peptides, and it is beyond the scope of this chapter to review them all. The main purpose of the ensuing sections is to highlight how sample preparation of body fluids affects the downstream separation procedures and MS. To this end, two of the major separation methods will be highlighted, notably 2D electrophoresis for whole proteins and HPLC for the low-molecular-weight fraction of proteins and peptides or for protein digests.

3.1. 2D Gel Electrophoresis of Proteins in Body Fluids

The presence of a few high-abundance proteins in most body fluids poses a problem for 2D electrophoresis. 2D gels have a limited loading capacity of some hundred micrograms of total protein, which makes the detection of medium- to low-abundance proteins difficult if not impossible unless the sample is prefractionated. Considering that albumin in serum represents 40 mg/mL of the 80 mg/mL total protein concentration and that established tumor markers circulate at concentrations of a few ng/mL or even less, it is clear that applying, for example, 500 μg of total protein to a gel (corresponding to approximately 6 μL of serum) will yield only about 10 pg (approximately 0.12 fmol for a protein of 50 kDa) of a given tumor marker in the original sample. Even assuming a recovery of 100%, this is clearly below the detection level of any protein staining technique and definitely an amount that cannot be identified by in-gel digestion and MS. Without any enrichment or prefractionation, 2D electrophoresis will not be able to reveal proteins at concentrations much below the μg/mL range, an area that is largely occupied by well-known plasma or serum proteins that are likely not relevant as disease-specific biomarkers (Figure 3) [8,133].

One way to enhance the capacity of 2D electrophoresis to detect proteins at lower concentrations is to remove high-abundance proteins selectively. An affinity column developed to deplete the six most abundant proteins from serum resulted in a reduction in total protein by about a factor of 10 [50]. The effect of this removal step on 2D electrophoresis can be appreciated in Fig. 7, which shows that after depletion a range of proteins become visible. However, increasing the loading capacity by a factor of 10 is not sufficient to reach into the ng/mL concentration range for complex body fluids like serum.

Another strategy to cover more of the low-abundance proteins is to enrich some of them specifically. This is naturally at the expense of losing the overview over the proteome and thus potentially missing relevant markers. Enrichment depends critically on the selection of appropriate ligands in
Introduction

combination with stationary phases of low nonspecific protein binding. Both requirements are not easily met but it is often the elimination of nonspecific binding that poses the greatest problems. As an example of the effect of nonspecific binding, Figure 9 shows the binding kinetics of a metalloprotease to controlled porosity glass beads (with or without silanization to render the surface more hydrophilic) and to the silanized beads containing an immobilized metalloprotease inhibitor.

Prefractionation of the sample is another option to reduce complexity prior to 2D electrophoresis, at the expense, however, of having to run multiple gels for a single sample. This is often not a viable option owing to the work-intensive nature of 2D gels. An alternative is to select narrow pH ranges to visualize only that part of the proteome that does not coincide with the high-abundance proteins. Unfortunately, most serum proteins have similar isoelectric points between pH 5 and 6 (see Figure 7), making fractionation difficult. More recently, prefractionation by preparative in-solution isoelectric focusing has emerged as a first step in body fluid analysis prior to 2D electrophoresis and also chromatography [134]. This approach has the advantage that proteins are fractionated based on a clear-cut physicochemical parameter, their isoelectric point, but multiple fractions still need to be analyzed, meaning that an efficient, preferably automated method should be used downstream [135]. The possibilities of prefractionating [136] proteins in depleted serum/plasma on a newly developed “Macroporous Reversed-Phase C18 High-Recovery Protein Fractionation HPLC column (mRP)” prior digestion and analysis (Agilent, www.agilent.com/chem, Agilent technologies publication 5989-250EN) will be described in chapter VI.

3.2. LC-MS of Proteins and Peptides in Body Fluids

Based on the above discussion, it is obvious that 2D electrophoresis is not the method of choice for analyzing large series of clinical samples in quest of new disease-specific markers. Consequently, other methods have been sought to reduce the workload of 2D gels, methods that make use of automated equipment. In addition to the already described direct combination of sample preparation on protein chips or magnetic beads with MS, there is increasing interest in the combination of online sample preparation with LC (LC-MS). In the following two examples, a focus on the low-molecular weight part of the body fluid proteome (also referred to as the peptidome) and the shotgun proteomics approach requiring proteolytic digestion will be highlighted. Indeed, there are many possibilities of integrating sample preparation with the ensuing separation, but these two approaches may serve as examples.

3.2.1. Peptidomics

Dividing the proteome of body fluids into a high- and a low-molecular weight fraction (the so-called peptidome [14,137-140]) is an approach to detect
lower abundance small proteins and peptides. Although it is restricted to a certain molecular weight range, the peptidome contains extensive information about processes in the organism that may be relevant for diagnosis and follow-up of therapy. This is a deliberate choice of sample pretreatment, eliminating most of the high-abundance serum or plasma proteins. An additional advantage of focusing on the molecular weight region below 15 to 20 kDa is that these molecules are more easily separated and recovered by reversed-phase HPLC (RP-HPLC), which is the preferred method for coupling to MS.

There are a number of techniques that allow elimination of the fraction of the proteome above approximately 20 kDa, such as ultrafiltration, precipitation with acids or organic solvents or the combination of ultrafiltration with adsorption chromatography (e.g., RAM chromatography) [15,16]. Full integration of all analytical steps in an automated system is often desirable for biofluid analysis in a clinical or biomedical environment to increase throughput, reduce the need for skilled personnel, and increase reproducibility. Furthermore, documentation is often facilitated by using an integrated, fully automated analytical system. Combining the “unit operations” of sample pretreatment, separation, and detection in the case of peptidomics was achieved in a system described by Wagner et al. [17] and further developed by Machtejevas et al. [18]. Figure 11 shows the instrumental setup combining selection of the peptidome from human hemofiltrate by RAM chromatography followed by prefractionation on a strong cation exchanger and finally separation by RP-HPLC.

Although this setup was not coupled online to a mass spectrometer, analysis of selected fractions after RP-HPLC by MALDI-TOF-MS showed that complexity of the original hemofiltrate had been reduced to such a level that most of the fractions contained one major peptide or small protein (Figures 12 and 13).
Introduction

Figure 12. Selected reversed-phase chromatograms from human hemofiltrate processed through the integrated, multidimensional chromatography system shown in Fig. 11. Numbered and marked peak fractions 1 to 10 were selected for MS analysis. Reproduced with permission from Machtejevas E, John H, Wagner K, et al. Automated multi-dimensional liquid chromatography: sample preparation and identification of peptides from human blood filtrate. *J Chromatogr B* 2004;803:121-130.
Figure 13. MALDI-TOF mass spectra of selected peaks from the reversed-phase HPLC fractions shown in Fig. 12. Spectra (A), (B), (C), and (D) correspond to peak fractions 4, 7, 8, and 9, respectively. Measurement was performed in the linear mode with positive ionization using a matrix of α-cyano-4-hydroxycinnamic acid mixed with fucose. The peak at \( m/z = 3914.4 \) is an internal standard. Reproduced with permission from Machtejevas E, John H, Wagner K, et al. Automated multi-dimensional liquid chromatography: sample preparation and identification of peptides from human blood filtrate. *J Chromatogr B* 2004;803:121–130.

### 3.2.2. Shotgun Proteomics

Applying HPLC separation to the high-molecular-weight region of the proteome requires prior proteolytic digestion, since most complex proteins are not stable under the denaturing conditions of RP-HPLC and are thus not quantitatively recovered. In addition, it is not possible to identify proteins based on their molecular weight only, and fragmentation of large proteins is impossible in most commonly employed mass spectrometers.

This has led to the shotgun proteomics approach, whereby the entire proteome is first digested with trypsin followed by one- or two-dimensional chromatographic separations of the peptides [20-23]. Most of the observations about high-abundance proteins masking those of lower abundance that were
made with respect to 2D electrophoresis also apply to the shotgun method. This is partially because of the limited loading capacity of the chromatographic columns, especially when nanoLC-MS (loadability in the ng–μg range) is the final analytical step. Removing high-abundance proteins by affinity chromatography prior to digestion is one way of increasing the loading of medium- and low-abundance proteins. The effect of depletion of albumin on the detection of a tryptic peptide derived from α1-antitrypsin, whose concentration in serum is about 20-fold lower, shows the clear improvement even for a protein that is still considered to belong to the high-abundance class (Fig. 6). Depleting high-abundance proteins and notably albumin can also have drawbacks, however, as albumin is known to bind and carry numerous compounds including small proteins and peptides that may be co-depleted. Zhou et al. [37] showed, for example, that albumin binds some 210 proteins and peptides, which could be identified in the albumin-containing fraction after depletion. In some of our own studies, we observed that a small protein added as an internal standard (horse heart Cytochrome C) at pmol/μL concentrations was co-depleted to leave about 25% [61] of the original amount when the six most abundant proteins were removed by affinity chromatography. It is important to note that although depletion removes some other proteins and peptides, this seems to be rather reproducible, at least when judged from the results with horse heart Cytochrome C (the relative standard deviation of peak areas without normalization is 10–30%) and our recent data with a large series of serum samples from prostate cancer patients [28].

The enormously complex mixture generated by shotgun proteomics usually requires multiple separation dimensions prior to MS. This has inspired researchers to develop novel ways of performing “coupled column” HPLC multidimensional protein identification technology (MudPIT) [21] or the integrated setup outlined in Fig. 11. Because analysis of individual fractions from earlier dimensions in such a multidimensional HPLC approach requires 12 to 24 h per sample, more efficient and rapid separation methods are under development. One of them is based on reducing the particle diameter of the chromatographic materials to 1-2 μm, which results in rather high backpressures that can only be delivered by special HPLC equipment [141-144]. Other approaches are based on monolithic materials that support very high linear flow rates at pressures amenable to common HPLC equipment [24,145-149] or the use of pellicular stationary phases also at elevated temperatures [150-153]. These developments show that in addition to the major advances in MS, which have made proteomics as we know it today possible, there is also considerable activity in the fields of sample preparation and separation methodology. It is only through integration of these unit operations into an analytical strategy that the challenges of body fluid analysis can be tackled and possibly the entire dynamic range covered. Much remains to be done.
4. Conclusions

Proteomics of body fluids is a rapidly expanding field driven by the search for better biomarkers for disease diagnosis, follow-up on therapy and evaluation of the response of patients to newly developed pharmaceuticals. The analysis of body fluids has a long tradition in clinical chemistry and serves to support decision making by clinicians in many respects. Because of recent methodological developments in separation science, MS, and bioinformatics, there has been a surge of efforts to apply them to biomarker discovery, often focusing on biomarker patterns rather than individual molecules. Sample preparation, the indispensable and very critical first step in an analytical method, has attracted less attention, and its relevance is often underestimated. As outlined in this chapter, the approach to sample preparation is an important decision of strategic relevance for the ensuing analyses. It is therefore pivotal to weigh the pros and cons of each approach in light of the final goal. We hope that the overview given in this chapter will guide the reader in this complex methodological field.

Acknowledgments

We thank Dr. Theo Reijmers (Leiden University, Division of Analytical Biosciences) for calculations related to Fig. 2 and Dr. Begona Barroso (Astellas BV, Leiderdorp) for contributing the data for Fig. 10. We are also very grateful to Dr. Peter Schulz-Knappe (BioVision, Hannover, Germany) for communicating unpublished information for use in this chapter.
References

Introduction


Introduction


Chapter I


Introduction

[128]. Kemperman RF, Horvatovich PL, Hoekman B, Reijmers TH, Muskiet FA, Bischoff R. Comparative Urine Analysis by Liquid Chromatography-Mass Spectrometry and
Chapter I


Introduction


