

## Review

# Approaching clinical proteomics: current state and future fields of application in fluid proteomics

Consensus document of the <sup>a</sup>DGKL – Deutsche Vereinte Gesellschaft für Klinische Chemie und Laboratoriumsmedizin ([www.dgkl.de](http://www.dgkl.de));

<sup>b</sup>DGPF – Deutsche Gesellschaft für Proteomforschung ([www.dgpf.org](http://www.dgpf.org));

<sup>c</sup>DGfZ – Deutsche Gesellschaft für Zytometrie ([www.dgfz.org](http://www.dgfz.org));

<sup>d</sup>SSCC – Schweizerische Gesellschaft für Klinische Chemie ([www.sccc.ch](http://www.sccc.ch));

<sup>e</sup>ÖGLMKC – Österreichische Gesellschaft für Laboratoriumsmedizin und Klinische Chemie ([www.oeglmkc.at](http://www.oeglmkc.at)); <sup>f</sup>Danubian Biobank Consortium ([www.danubianbiobank.de](http://www.danubianbiobank.de))

**Rolf Apweiler<sup>1,b</sup>, Charalampos Aslanidis<sup>2</sup>, Thomas Deufel<sup>3,a</sup>, Andreas Gerstner<sup>4,c</sup>, Jens Hansen<sup>2</sup>, Dennis Hochstrasser<sup>5,d</sup>, Roland Kellner<sup>6,b</sup>, Markus Kubicek<sup>7,e</sup>, Friedrich Lottspeich<sup>8,b</sup>, Edmund Maser<sup>9</sup>, Hans-Werner Mewes<sup>10</sup>, Helmut E. Meyer<sup>11,b</sup>, Stefan Müllner<sup>12,b,f</sup>, Wolfgang Mutter<sup>13,b</sup>, Michael Neumaier<sup>14,a</sup>, Peter Nollau<sup>15,a</sup>, Hans G. Nothwang<sup>16,a</sup>, Fredrik Ponten<sup>17</sup>, Andreas Radbruch<sup>18,c</sup>, Knut Reinert<sup>19</sup>, Gregor Rothe<sup>20,a</sup>, Hannes Stockinger<sup>21,c,f</sup>, Attila Tarnok<sup>22</sup>, Mike J. Taussig<sup>23</sup>, Andreas Thiel<sup>18,c</sup>, Joachim Thiery<sup>24,a</sup>, Marius Ueffing<sup>25,b</sup>, Günther Valet<sup>8,c</sup>, Joel Vandekerckhove<sup>26</sup>, Wiltrud Verhuvén<sup>27</sup>, Christoph Wagener<sup>15,a</sup>, Oswald Wagner<sup>7,e,f</sup> and Gerd Schmitz<sup>2,a,b,f,\*</sup>**

<sup>1</sup> European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, UK

<sup>2</sup> Institute of Clinical Chemistry and Laboratory Medicine, University of Regensburg, Regensburg, Germany

<sup>3</sup> Institute for Clinical Chemistry and Laboratory Diagnostics, University of Jena, Jena, Germany

<sup>4</sup> Department of Otorhinolaryngology/Surgery, University of Bonn, Bonn, Germany

<sup>5</sup> Department of Pathology/Clinical Chemistry, University of Geneva, Geneva, Switzerland

<sup>6</sup> Merck, Darmstadt, Germany

<sup>7</sup> Institute of Medical and Chemical Laboratory Diagnostics, University of Vienna, Vienna, Austria

<sup>8</sup> Max-Planck-Institute for Biochemistry, Martinsried/Munich, Germany

<sup>9</sup> Institute of Toxicology and Pharmacology, University Kiel, Kiel, Germany

<sup>10</sup> Institute for Bioinformatics, GSF, Neuherberg, Germany

<sup>11</sup> Medical Proteomics Center, University of Bochum, Bochum, Germany

<sup>12</sup> Protagen, Dortmund, Germany

<sup>13</sup> PROFOS, Regensburg, Germany

<sup>14</sup> Institute for Clinical Chemistry, University Hospital Mannheim, Mannheim, Germany

<sup>15</sup> Institute for Clinical Chemistry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>16</sup> Department of Animal Physiology, Technical University Kaiserslautern, Kaiserslautern, Germany

<sup>17</sup> Institute for Genetics and Pathology, University of Uppsala, Uppsala, Sweden

<sup>18</sup> German Rheumatism Research Center Berlin, Berlin, Germany

<sup>19</sup> Department of Mathematics and Computer Science, University of Berlin, Berlin, Germany

<sup>20</sup> Laboratory Center Bremen, Bremen, Germany

<sup>21</sup> Department of Molecular Immunology, University of Vienna, Vienna, Austria

<sup>22</sup> Department of Pediatric Cardiology, University of Leipzig, Leipzig, Germany

<sup>23</sup> Technology Research Group, The Babraham Institute Cambridge, Cambridge, UK

<sup>24</sup> Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany

<sup>25</sup> Institute of Human Genetics, GSF Neuherberg, Neuherberg, Germany

<sup>26</sup> Department of Medical Protein Research, University of Gent, Gent, Belgium

<sup>27</sup> Waters, Eschborn, Germany

\*Corresponding author: Prof. Dr. med. Gerd Schmitz, Institute of Clinical Chemistry and Laboratory Medicine, University of Regensburg, Franz-Josef-Strauß Allee 11, 93053 Regensburg, Germany  
Phone: +49-941-944-6201, Fax: +49-941-944-6202,  
E-mail: [gerd.schmitz@klinik.uni-regensburg.de](mailto:gerd.schmitz@klinik.uni-regensburg.de)  
Received October 14, 2008; accepted May 12, 2009

## Abstract

The field of clinical proteomics offers opportunities to identify new disease biomarkers in body fluids, cells and tissues. These biomarkers can be used in clinical

applications for diagnosis, stratification of patients for specific treatment, or therapy monitoring. New protein array formats and improved spectrometry technologies have brought these analyses to a level with potential for use in clinical diagnostics. The nature of the human body fluid proteome with its large dynamic range of protein concentrations presents problems with quantitation. The extreme complexity of the proteome in body fluids presents enormous challenges and requires the establishment of standard operating procedures for handling of specimens, increasing sensitivity for detection and bioinformatical tools for distribution of proteomic data into the public domain. From studies of *in vitro* diagnostics, especially in clinical chemistry, it is evident that most errors occur in the preanalytical phase and during implementation of the diagnostic strategy. This is also true for clinical proteomics, and especially for fluid proteomics because of the multiple pretreatment processes. These processes include depletion of high-abundance proteins from plasma or enrichment processes for urine where biological variation or differences in proteolytic activities in the sample along with preanalytical variables such as inter- and intra-assay variability will likely influence the results of proteomics studies. However, before proteomic analysis can be introduced at a broader level into the clinical setting, standardization of the preanalytical phase including patient preparation, sample collection, sample preparation, sample storage, measurement and data analysis needs to be improved. In this review, we discuss the recent technological advances and applications that fulfil the criteria for clinical proteomics, with the focus on fluid proteomics. These advances relate to preanalytical factors, analytical standardization and quality-control measures required for effective implementation into routine laboratory testing in order to generate clinically useful information. With new disease biomarker candidates, it will be crucial to design and perform clinical studies that can identify novel diagnostic strategies based on these techniques, and to validate their impact on clinical decision-making. Clin Chem Lab Med 2009;47:724–44.

**Keywords:** cerebrospinal fluid (CSF); clinical proteomics; fluid proteomics; mass spectrometry (MS); matrix assisted laser desorption/ionization (MALDI); preanalytical effects; standard operating procedures (SOP); surface-enhanced laser desorption/ionization (SELDI).

## Introduction

The proteome of an organism, as the complement of its genome, is highly dynamic and varies according to cell type and functional state. These effects in protein composition may be observed in body fluids and may reflect immediate and characteristic changes in response to disease processes and external stimulation. Clinical proteomics is the field that encompasses the quantitative and qualitative profiling of proteins

and peptides that are present in clinical specimens like tissues and body fluids. The actual proteome present in body fluids, cells and tissues at a certain point of time cannot be directly predicted from genomic information because, it represents only a subset of all possible gene products. Proteins may exist in multiple forms within cells or between various cells due to post-translational modifications (PTMs) or degradation processes that affect protein structure, function, localization and turnover. It is possible that in addition to the proteome, proteolytic degradation products, termed low-molecular-weight (LMW) range proteome, also may contain disease-specific information and lead to the identification of disease-specific biomarkers. Disease-specific peptides often appear as fragments of endogenous high-abundant proteins (e.g., transthyretin), or fragments of low-abundance cellular and tissue proteins, such as BRCA-2 (breast cancer) (1, 2). In addition, the extent of degradation of proteins by catabolic pathways may depend on preanalytical variables like temperature or time of storage of the sample. Therefore, special care must be given to specimen handling.

Our ability to identify and characterize molecules for early detection of disease or stratification of disease and to expand the prognostic capability of current proteomic modalities is enhanced by emerging novel nanotechnology strategies that make use of these LMW biomarkers *in vivo* or *ex vivo*. However, they may also lead to new problems related to accuracy and variation. Pathophysiological processes that involve proteolytic activities such as tumor proteases, are detectable in the plasma peptidome (1). The identification and characterization of highly specific and sensitive proteins or biomarker panels for risk stratification, prognostic assessment or early detection of disease is the key to treatment of complex diseases such as cancer, neurodegenerative disorders and metabolic and vascular disease.

The focus of clinical proteomics is on the analytical and clinical validation and implementation of novel diagnostic or therapy related markers identified in preclinical studies such as, for example, drug screening studies. Clinical proteomics, with an emphasis on fluid proteomics, also includes the selection, validation, and assessment of standard operating procedures (SOPs) in order that adequate and robust methods are integrated into the workflow of clinical laboratories. Standard measures need to be introduced in order to protect specimens from cell lysis, non-specific proteolysis and modification during collection, transport, and preparation prior to analysis. Useful considerations of preanalytical variables concerning body fluids are reported in the review of Paik et al. (3).

Selection of potential targets for clinical proteomics follows either a top-down or a bottom-up approach. The top-down approach applies high-throughput technologies to either population-based studies or selected cohorts, such as case/control studies or twin registers, in order to identify novel markers using an unbiased approach. These markers have to be further

characterized with respect to sensitivity, specificity, and function. The bottom-up approach focuses on previously identified pathways and looks for protein-protein or metabolite interactions, or interactions related to the pathway.

Recent advances in proteomic analysis due to use of high-throughput and high-content analysis has paved the way for clinical proteomics. These advances have been realized in the field of affinity binder technologies and in liquid phase technologies such as tandem mass spectrometry (MS/MS). For fluid proteomics, Luque-Garcia and Neubert have reviewed and summarized sample preparation for profiling and biomarker identification using mass spectrometry (MS) (4). High-resolution liquid phase separation methods, along with advances in chemometry and biometry for large-scale data analysis, now, offer the possibility for introducing these tools into laboratory diagnostics. This will enable screening for risk factors, identification of new disease-specific or stage-specific biomarkers and identification of novel markers for therapeutic drug monitoring or new therapeutic targets. Clinical proteomics has the potential to complement genomics, metabolomics, lipidomics, glycomics and transcriptomics, including splice variant analysis, and to contribute to a better understanding of disease processes. This will enable translation of this complex knowledge into diagnostic tools for clinicians.

Clinical proteomics is now on the verge of entering the hospital, similar to the field of metabolomics, which has now been established for clinical diagnosis in newborn screening. However, before this occurs, essential criteria for successful use in a clinical environment need to be fulfilled. First, it is crucial that high-throughput analytical platforms be implemented that provide reproducible protein patterns with a clinically acceptable turnaround time. Also, they need to be robust enough for everyday use, simple enough to be operated by technicians with a minimum of supervision required, and be able to fit into the clinical laboratory workflow. Second, bioinformatic algorithms need to be developed that include chemometry, data reduction and conversion into actionable health information. These algorithms also must be robust and easily integrated into current laboratory information systems. Third, the preanalytical conditions for clinical specimens need to be standardized and optimized for the development of clinically applicable tests. Prior to research on the development of novel biomarkers, appropriate and well-defined patient cohorts that address a specific clinical question need to be selected. These cohorts should be well-characterized by appropriate anamnestic and physiologic parameters including age, sex, hormonal status, treatment and hospitalization status. This information should be made available for study. Furthermore, SOP-driven biobanks and biorepository systems have to be established and integrated into the diagnostic workflow and storage conditions should be validated. Undoubtedly standardized and SOP-driven preparation of patient samples for body fluid proteomics, like any other clinical laboratory

analysis, is one of the most urgent challenges for obtaining reproducible and clinically useful results. Furthermore, regulations must be established to address medico-legal issues such as patient consent and commercial use of samples, as well as intellectual property. This must be achieved at a supra-national level to allow for large-scale multi center studies, which are a prerequisite for the task at hand.

In this review, we focus on fluid proteomics and suggest procedures for sample preparation and standardization of protocols for analysis of body fluids. We summarize the preparative and analytical methods in the field and emphasize key clinical applications. We conclude with bioinformatic approaches in proteomics. Another review focusing on cellular proteomics (cytomics) is being prepared as a consensus document from the authors and will follow.

### Preanalytical phase

It is well known that sources of technical or sample variation are found primarily in the preanalytical phase. Lack of standardized procedures for patient preparation (e.g., fasting, diurnal rhythm), specimen acquisition, handling, and storage account for more than 90% of the errors within the entire diagnostic process (5). Advances in genomics and proteomics have led to high expectations for clinical biomarker discovery and use. For successful generation of validated biomarkers, more attention has to be focused on the preanalytical stage in areas such as sample collection, transport, preparation, and processing (6). In addition, standardization and quality management procedures, in particular when large biorepositories (biobanks) are being established and used, need to be addressed. The preclinical discovery phase will target promising biomarkers that require validation before translation into clinical proteomics can occur. The type of sample needed, as well as sample processing, could be quite different for the different phases of biomarker development and biomarker validation. There is a considerable difference between the requirements for high-throughput proteomic profiling for clinical proteomics, and subsequent protein identification and in-depth characterization of single protein samples (low-throughput) in the preclinical and discovery phase. It is well known that individual clinical chemistry and hematology parameters are prone to significant influence by preanalytical handling of biological fluids. It is very likely that proteomic profiles consisting of a plethora of individual parameters will also be susceptible to preanalytical handling.

Among the plethora of possible variables from which preanalytical effects can arise are the site of sample collection, the process of blood collection, the material and liquid content of the sample container, the time until sample processing and the temperature (7).

Cerebrospinal fluid (CSF) parameters such as protein composition and total protein content vary significantly depending of the site of sample collection,

as for example from the ventricles obtained during surgery or by lumbar puncture. Also, the site used for collection of blood samples is important (8). In vitro hemolysis causing leakage of hemoglobin and other intracellular erythrocyte components into the blood fluid can occur when blood is collected from fragile veins such as hand veins instead of veins in the antecubital area. Hemoglobin interferes with a variety of analytical reactions and causes an increase in absorbance measurements, especially at 415, 540 and 570 nm (9–15). Even if hemolysis is not visible, hemoglobin chains can be detected on polyacrylamide gels and occur as characteristic peaks on matrix assisted laser desorption/ionization (MALDI) or surface-enhanced laser desorption/ionization (SELDI) spectra (16, 17). It can be concluded that hemolysis might have a significant impact on identification of proteomic parameters and, therefore, should be avoided. Another intracellular molecule in erythrocytes that can interfere with laboratory tests is adenylate kinase. This molecule interferes with the determination of creatin kinase and causes falsely increased results.

The time of sample collection is also a preanalytic factor not to be underestimated (18). Many parameters such as peptide hormones and cytokines have diurnal rhythms. For the collection of urine, consideration must also be given to the time that urine is stored in the urinary bladder prior to collection. It has been shown that first void morning urine is more prone to protein degradation because of bacterial contamination, compared with urine collected at other times (19). In addition, significant differences in proteomic profiles of urine determined by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) were documented to be dependent on the collection time (20). The greatest differences seem to occur between urine collected at midday and the first and second morning urine. One should also keep in mind the importance of patient preparation as, for example, parameters that depend on metabolic status where fasting before sample collection is essential.

Only a few studies have been performed that demonstrate the effects that sample collection procedures have on proteomic profiling; handling and storage had comparably lower effects (16).

Another important issue that impacts determination of certain parameters is the mode of specimen collection. Needle bore size, patient posture and tourniquet application have a significant effect on parameters such as total protein, albumin, IgGs or erythrocyte integrity (21–23). The type of sample container influences parameter results. Serum and plasma differ as a consequence of clotting. In addition to the reduction in proteins and fibrinogen due to the clotting cascade, many other differences have been documented. Generally there are more peptides in serum, but the type of anticoagulant also influences the peptide composition (16, 24–27). Complement C3f, for example, is released by factors I and H and is a fragment of complement C3b, a cleavage product of complement C3 (28). Complement C3f was found in higher concentra-

tions in serum than plasma (29), indicating that the full-length peptide is generated by a post-coagulation exoprotease. The clotting time also has a significant influence on parameter results. It has been observed that the intensity of protein peaks determined by SELDI-TOF changes significantly during clotting of serum from 30 min to 60 min (30). After 60 min, changes could only be documented if the sample was stored at room temperature, but not if the sample was stored on ice. This study indicates that serum should be prepared 60 min following clotting at room temperature, followed by storage at 4°C or even better at –80°C before analysis.

Tube and anticoagulant type can also influence sample processing. EDTA chelates divalent cations and, therefore, is not suitable for assays dependent on such ions. Highly charged heparin molecules may interact with proteins and alter their separation characteristics when using chromatography (27). If heparin is used as the anticoagulant, dilution effects have to be taken into consideration, especially if the tube is not filled completely. Tubes containing a gel-based separator can alter sample composition. It is not known whether this alteration is caused by effects on the clotting cascade or by effects of the gel itself on the proteins (16, 27). For  $\beta$ -amyloid and  $\tau$ -protein determination in CSF, the tube material has to be chosen carefully. CSF should not be collected in polystyrene, but rather in polypropylene tubes (31) to prevent erroneous results. On the other hand, substances released from the material, e.g., polyvinylpyrrolidone (PVP), can have a direct impact on MALDI-TOF spectra by generating multiple interfering peaks in the  $m/z$  range of 1000–3000 (32). Polymers from different brands of tubes and heparin have been described as inducing ion suppression effects (33).

Processing temperature, elapsed time following venipuncture before separation of plasma or serum from cells and centrifugation speed and duration may have a significant impact on parameter results. The duration between post-venipuncture and sample processing may cause different stages of coagulation and complement processes, or the release of cell-derived products, and influence parameter results (16, 24, 25, 34).

The effects of temperature are ambiguous. Low temperature minimizes proteolytic activity. However, blood samples also contain various antiproteases, therefore, cooling may not be necessary. In addition, for some analytes such as platelets, it is recommended that samples are processed at 18–20°C to prevent platelet activation. This could also have an effect on the plasma proteome, especially, for small proteins. Erythrocytes are less stable at lower temperatures. Therefore, release of intra-erythrocytic proteins into the plasma is more likely due to hemolysis.

Proteolysis can also continue in frozen samples. Differences between samples frozen for 1–3 months at –20°C or –80°C have not been observed in several studies (16, 26, 34). However, this may be due to the short storage times because another research group reported that most differences became apparent after



about 5 months of storage (35). One peptide noted to be altered by storage at low temperature was a biomarker previously identified as a candidate for colorectal cancer, an N-terminal albumin fragment (m/z 3087). This underlies the importance of correct preanalytical handling to prevent false results. Another protein influenced by storage conditions is complement C3f, discussed above. The corresponding peak was seen in samples frozen at  $-20^{\circ}\text{C}$  for  $<1$  month, but not in samples stored at  $-70^{\circ}\text{C}$  (35). Since there was a higher C3f concentration when stored at  $-20^{\circ}\text{C}$ , it is likely that the exoprotease continues to generate C3f at low temperature storage conditions. In addition to serum and plasma proteins and peptides being affected by storage, proteins of diagnostic relevance collected from CSF are also affected. It has been shown that cystatin C, a LMW cysteine proteinase inhibitor, is cleaved with the resulting loss of its N-terminal amino-acids when stored for 3 months at  $-20^{\circ}\text{C}$ , but not at  $-70^{\circ}\text{C}$  (36). If not considered, such issues can lead to incorrect conclusions. A recent study using SELDI-TOF on CSF from patients with multiple sclerosis proposed that a unique 12.5 kDa protein peak, reportedly a C-terminally cleaved cystatin C isoform, enabled distinction of multiple sclerosis from other neurological diseases with 100% specificity (37). However, a later study showed that this result was a storage related artefact and not useful as a diagnostic marker for multiple sclerosis (38). The 12.5 kDa cystatin C isoform was produced from full-length cystatin C during storage at  $-20^{\circ}\text{C}$ .

However, proteolytic cleavage that occurs following blood collection could also reflect disease-specific protease activities, and, therefore be the result of a particular disease such as cancer (29). It was shown that fragments observed after sample storage at  $-20^{\circ}\text{C}$  still allow discrimination of colorectal cancer patients and healthy controls (35).

Another important issue that is very likely to alter parameter results are freeze-thaw cycles, although the effects have not been investigated systematically by many groups. In one systematic investigation, sera of eight patients with sarcoidosis and eight controls were frozen and then thawed between one and eight times and then spotted on a CM10 (cation exchange) and on a NP20 (normal phase) ProteinChip array (39). Three different peaks distinguishing patients and controls could be identified in the samples frozen and then thawed more frequently. In contrast, using freshly frozen sera, none of these markers, except for another significant single peak, was identified. These results indicate the importance of avoiding freeze-thaw cycles.

To address preanalytical effects on parameter results, standard protocols for serum and plasma sampling, handling and storage are required. This necessity does not arise from the issue of which procedure is better, but rather standardized procedures need to be used in order to obtain comparable and reproducible results between different laboratories or research groups (26).

### Standardization of the (pre)analytical process

To achieve reproducible clinical proteomics results, standardization is an essential requirement. The importance of standardization is demonstrated by three different studies on prostate cancer which found completely different decision trees using identical chip types and comparable study populations (40–42).

Examples of systematic bias errors resulting in false-positive and false-negative results include 1) preanalytical variables such as systematic differences in study populations and/or sample collection, handling, and pre-processing procedures; 2) within-class biological variability which may comprise unknown sub-phenotypes among study populations; 3) analytical variables such as inconsistencies in instrument conditions and reagents which result in poor reproducibility, and 4) measurement imprecision (18, 43).

To minimize these effects and to allow a comparison among different clinical studies, researchers and clinicians need to standardize and clearly describe the protocols used and the performance of clinical studies. The study aims have to be clearly defined, patients and controls have to be accurately characterized, the samples that are collected, the collection procedures, the sample preparation and processing methods and data modeling algorithms have to be carefully documented. For example, with respect to the characterization of patients and controls, one interesting question is whether controls should consist not only of healthy individuals, but also of patients suffering from another disease affecting the same organ or producing the same symptoms. The latter case is clinically more interesting because it aims to identify biomarker panels suited for differential diagnosis instead for the identification of sick individuals. But for this approach a detailed description of the controls is needed, rather than just referring to them as the control group. Existing standardization protocols with good reputation, for example the standard of "good clinical practice" (GCP) (See European Union Directive 2001/20/EC) and the "good clinical laboratory practice" (GCLP) (7) should be utilized for clinical proteomic studies.

The effects of preanalytical variation and the need to standardize the preanalytical phase is highlighted by Marshall et al. (44). Using MALDI-TOF-MS, these authors analyzed blood samples obtained from patients with myocardial infarction and found that recorded changes in the protein profiles corresponded to serum protease activities, rather than being the result of disease processes.

Following identification of a single biomarker or a biomarker panel, a clear definition of the protein(s), including sequence and PTMs is mandatory as part of the standardization process. However, since many standard methods such as those for MS/MS applications do not consider PTMs, it is necessary to accurately document the physico-chemical properties necessary for detection of the identified biomarker. The physico-chemical properties also have to be doc-

umented since they can significantly influence the results. For example, they can influence the dynamic range and sensitivity of the applied method as demonstrated by the stains used on protein gels and their differences in dynamic range and sensitivity of protein quantification. For instance, in contrast to the older Coomassie blue and silver stains, the newer fluorescent stains have a higher dynamic range and a similar sensitivity (45).

With quality control, standardization should also include the acceptable deviation for identical analytes detected in different samples. Proteome data are always prone to error, but the extent needs to be detected and thresholds defined.

Standardization does not end after a study is performed and data collected, but should also include data storage. For multivariate analyses like proteomics, a huge amount of data is necessary. To allow the comparison and further development of different proteomic investigations it is necessary to store proteomic data in a standardized data format using very exact and precise rules. Apart from the final results, this database should also include all information about the study protocol, patient and control phenotyping, sample collection and processing, quality control criteria, calibration and matching.

Another source of variation to be considered in biomarker discovery is the heterogeneity of the human population (biological variation), referred to as between-subject variation. Between-subject variation is the sum of differences in protein expression between people resulting in, but not limited to, differences in age, gender, or race (46). Of course, this kind of biological variation cannot be avoided but has to be taken as a fact and considered in the analytical proteomic strategy as well as in interpretation of results.

The use of control materials, e.g., the addition of standards, is mandatory for controlling the linear range, specificity, sensitivity, precision, and accuracy. Once the instrument is calibrated, standardization of the output can be accomplished by running control samples and adjusting parameters such as laser intensity, detector voltage, and detector sensitivity to ensure that spectra are consistent between runs. Controls, either as internal or added analytes, should also be analyzed throughout an experimental run; the former is preferable, if possible, as internal controls would not interfere at all with the analytical process. A recent study using SELDI-TOF on prostate cancer cases and control cases demonstrated that implementation of such strategies on SELDI-TOF approach may provide reproducible results for peak patterns (47). However, other issues for this approach need to be discussed further. Transition of MS technology from a research tool to a reliable clinical diagnostic platform requires rigorous chemometric standardization, spectral quality control and assurance, SOP for robotic and automatic sample application, and standardized controls to ensure the generation of highly reproducible spectra. The introduction of peptide

standards at defined quantities, as well as the use of quantitative labeling techniques (e.g., stable isotope labeling) for routine applications, may eventually allow quantitative assessment of selected markers. Currently, laboratories are independently developing their own methods, optimization procedures and in-process controls, but effort is lacking for standardizing methods between laboratories. Current instrumentation is at the level of "advanced prototypes" that provide research tools for specialized laboratories, but do not yet qualify for routine laboratory testing. However, the development of standard technologies for MS platforms with reference standards for controls and calibrators will certainly help to accelerate the process of evaluating proteomics technologies for clinical applications. This transition will require widespread collaborative efforts between public health organizations, legislation, industry, researchers, healthcare providers, health insurance companies and patient organizations.

With advances in analytical instrumentation and reagent quality, and the availability of high-quality analytical standards, analytical errors are no longer the primary factors influencing the reliability and clinical utilization of laboratory diagnostics. An example of a standardized protocol for sampling and preparation of blood as established by the National biobank program in Sweden can be found at <http://www.biobanks.se> (SOP – Collection of blood samples Gerd Johansson Göran Hallmans, Medicinska biobanken).

## Proteomics in body fluids

### Techniques for proteomic analysis

The techniques used for protein analysis can generally be divided into "unbiased" and "biased" methods. In "unbiased" techniques, the investigator does not preselect the proteins to be examined, but searches for changes in any proteins that are identified. Such methods typically include multiple protein separation techniques such as 2-dimensional sodium dodecylsulfate polyacrylamide gel electrophoresis (2D-PAGE) and SELDI-TOF. Protein subpopulations are enriched according to their physico-chemical or immunological characteristics by binding to a variety of modified chip surfaces prior to MALDI-TOF-MS ("SELDI-TOF" MS), liquid chromatography (LC) methods where the protein mix flows through a column packed with porous beads of particular binding properties, and capillary electrophoresis (CE) where proteins are separated based on their charge dependent migration in an electric field. These separation strategies are followed by protein identification using MALDI-TOF-TOF or electrospray ionization tandem mass spectrometry (ESI-MS-MS). "Biased" techniques include antibody-based affinity-binding methods or multiple reaction monitoring tandem mass spectrometry (MRM-MS-MS), where the proteins of interest have already been identified. Highly specific

antibodies against these proteins or specific peptides are synthesized and labeled. This allows for more in depth profiling of these preselected proteins, enabling protein abundance to increase by more than eight orders of magnitude.

**2D-PAGE** The first technique used for separation in proteomics is often 2D-PAGE. With this technique, proteins are separated both according to their isoelectric point (pI) and mass (combination of isoelectric focusing and SDS-PAGE). The resolution power of 2D-PAGE allows the identification of 1000–2000 protein spots. In theory, it is possible to visualize up to 10,000 protein spots in a single large gel (48). To achieve higher resolution and to improve the detection of low-abundance proteins, several 2D-GE can be used with overlapping narrow pH gradients which helps reduce the amount of protein/spot (49–51). Proteins can be visualized using Coomassie brilliant blue or silver nitrate, or by the use of a sample pre-labeled with fluorescent dyes such as SYPRO Ruby and cyanine dyes (Cy2, Cy3, Cy5) (40, 45, 52–54). The use of fluorescent dyes increases sensitivity, offers a linear dynamic range exceeding up to three orders of magnitude and allows for quantitative comparison of gel-based protein patterns (55, 56).

Post-translational protein modifications that alter the pI or the overall molecular mass can be identified by changes in the x/y position within the 2D gel. Since changes of a single charge are detectable with this technique, the effects of post-translational processing such as differently phosphorylated or glycosylated forms of the same protein can be identified as a series of spots having the same molecular weight (57–59). Spots of interest can be excised and subjected to proteolytic digestion. The resulting peptides are then analyzed by MALDI-TOF-MS or LC-MS and compared with theoretical spectra from databases to identify the protein. Quantitation of proteins in 2D-PAGE has been largely improved with the development of the difference gel electrophoresis (DIGE) technique (60). In DIGE-based proteomics, up to three samples are derivatized using different fluorophores of the cyanine series and then run on the same gel. This is currently one of the most reliable routine platforms for quantitative proteomics.

Its strengths are low experimental variation due to mixing of experimental and control samples, and inclusion of an internal standard, potential visualization of protein isoforms including splice variants and PTMs, and precise information obtained on molecular weight and pI (60, 61). In addition, the cost for 2D electrophoresis is relatively low, which is an important issue for healthcare systems. However, the technology has distinct limitations. Proteins with either low (<5000 Da) or high (>1,50,000 Da) molecular weight and certain physico-chemical properties (e.g., hydrophobic membrane proteins) are very difficult to separate and detect, or may not be detectable at all. Furthermore, low-abundant proteins in complex samples such as plasma or cell extracts are not detected because they are masked by highly abundant proteins

like albumin. This makes it impossible to determine low- and highly abundant proteins simultaneously. Sample fractionation, based, for example, on cellular components, can help overcome this limitation by reducing the complexity of the sample.

However, any method used to reduce complexity is bound to introduce artefacts and to increase noise, influencing the statistical significance of the data produced. High-throughput protein analysis is hampered by time consuming separation and tedious staining and destaining processes. The lack of widely accepted automation resulting in the requirement for experienced academic and technical staff make it difficult to introduce 2D-PAGE into the clinical routine. Until now, almost no parameter for clinical diagnostics is routinely measured using 2D-PAGE. This method is restricted to clinical research tasks such as discerning undefined fluids (skull fracture and suspected CSF leak), CSF (Creutzfeldt-Jakob disease marker), and plasma and urine analysis (undetectable paraproteinemia, etc.).

**Liquid chromatography (LC)** High-resolution LC separation coupled with MS has recently become a widely used platform for proteomics. The fractionation of samples is based on biophysical properties such as surface charge (ion exchange chromatography), hydrophobicity (hydrophobic interaction chromatography) or affinity for certain compounds [affinity, dye ligand, reversed-phase liquid chromatography (RPLC)] (49, 56, 62–64). These LC techniques can be applied to the characterization of complex protein mixtures utilizing two different strategies. The intact proteins can be separated and then digested prior to characterization by MS (“top-down” approach), or the complete mixture can be digested and the peptides separated and characterized (“bottom-up” or “shotgun” approach). Although the bottom-up strategy relies on the analysis of protein fragments that are sufficiently unique to enable identification of the parent protein, this strategy still accounts for the majority of proteomics approaches. Both strategies have advantages and disadvantages. While the chromatographic separation of certain proteins, such as those that are labile or hydrophobic, pose a serious problem to a top-down approach, some information (e.g., on splice variants, degradation, assignment of PTMs) may be lost in a bottom-up approach. Furthermore, digestion of a protein mixture (e.g., with trypsin) increases the complexity by at least one order of magnitude. In order to address this problem, multi-dimensional protein identification technologies (MudPIT) have been developed recently (65, 66). Following sample reduction, alkylation and digestion, the resulting peptide mixture is separated using cation-exchange chromatography followed by reverse-phase chromatography and tandem MS. With the recent development of high-speed 2D linear ion trap instruments such as linear trap quadrupole (LTQ), protein profiling coverage has been greatly enhanced compared with traditional three-dimensional ion trap systems (67).

In conclusion, LC-MS/MS technologies now routinely allow for the identification of thousands of proteins in complex samples from mammalian tissues and cells. Although routinely used for peptide/protein identification, data-dependent LC-MS/MS still has an inherent limitation of "undersampling", whereby only a portion of the species observed in the survey MS scan is selected for fragmentation (66, 68). The LC unbiased technique allows largely automated high-throughput analysis. However, quantitative data usually is not obtained in all laboratories. LC-MS/MS is already in use for routine clinical analysis, especially for inborn errors of metabolism. However, instruments dedicated to routine clinical use are still desired, and this has the potential to become a widely used method (see Applications).

**Capillary electrophoresis (CE)** CE, especially capillary zone electrophoresis (CZE), is frequently used as a front-end coupled to a mass spectrometer. The separation of proteins in CE is based primarily on their charge to mass ratios, size-sieving effects-, interaction with ligands (affinity) and/or hydrophobicity/hydrophilicity (partition). CE coupled to MS offers several advantages: 1) it provides fast separation and high-resolution (69), 2) it is robust and uses inexpensive capillaries instead of expensive LC columns (70), 3) it is compatible with most buffers and analytes (71), and 4) it provides a stable constant flow, thereby avoiding gradients in the buffer that may otherwise hamper MS detection (72). However, CE cannot be easily applied for high-molecular weight proteins, which can be removed effectively by techniques such as ultrafiltration (73). Another limitation of CE-MS is the relatively small sample volume that can be loaded onto the capillary, resulting potentially in lower sensitivity for detection, or loss of discrimination. Despite technical improvements, the complexities of human body fluids and the problem of high-dynamic range provides significant challenges for using CE for routine proteomic profiling.

In conclusion, although CE and CZE offer high-throughput and automation capabilities, along with sensitivity and good resolution power, the combination of these separation techniques with MS for routine proteomics has just recently become commercially available. This is due primarily to poor reproducibility, complicated configurations and the need of skilled operators (74).

**Mass spectrometry (MS)** MS technologies provide the backbone for the majority of proteomic research. The rise of proteomics can be attributed to the availability of completely sequenced genomes since 1995, together with advances in protein ionization, increased resolution, improved sensitivity and high-throughput MS analyzers available since the late 1980s, along with the development of protein computational search algorithms.

A mass spectrometer consists of an ion source to introduce the analytes to gas for ionization, one or more mass analyzers for measurement of the  $m/z$

ratio of the ionized analytes and a detector for the determination of the number of ions at every  $m/z$  ratio.

Of the ionization techniques available today, electrospray ionization (ESI), which forms the ions from a liquid solution, and matrix-assisted laser desorption/ionization (MALDI) using a laser pulse to sublimate the analyte from a dry matrix, provide high-sensitivity and can generate ions without significant chemical decomposition such as the breaking of covalent bonds (75–78). Using electrical voltage, ions are accelerated into the mass analyzer and the  $m/z$  values are measured by the motion of the ions through the mass analyzer. An example of a mass analyzer frequently used is the TOF analyzer that separates ions based on the differences in transit time (time of flight) from the ion source to the detector through tubes under vacuum. By placing two analyzers in sequence, two-stage MS can be performed (tandem MS or MS/MS). Peptide sequences are identified by comparing experimental mass spectra with theoretical mass spectra using protein databases and search algorithms such as SEQUEST, MASCOT, PHENYX and !X-TANDEM (79–83). The amino acid sequence of the whole protein can be determined by matching the mass spectrum to known spectra using search algorithms such as SEQUEST (81), MASCOT (80) or PHENYX (82).

One analytical problem is that ionization is never 100% efficient due to the physico-chemical properties of the analyte molecules, including pKa value, polarity, hydrophobic or hydrophilic index, and ionization potential. In addition to the variability in ionization potential between various peptides, the efficiency of ionization is directly influenced by the concentration and type of peptides infused into the atmospheric pressure ionization (API) source. The higher the peptide concentration, the lower the capacity to sufficiently ionise all available peptides due to depletion of all available protons and to upper mass density constraints in the dynamic range of the mass spectrometer.

The majority of MS protein identification is based either on the characterization of short, unique (~ 6–10 amino acid) peptide sequences (tags), or on spectra comparison. With the automation and software that is available, more than 1000 spots or bands can be prepared and measured by a single person per day using MALDI-TOF-TOF MS. A limitation of MALDI-TOF is the identification or detection of low molecular mass proteins which deliver so few peptides that identification is often based using a low number of matches. The MALDI-TOF-based protein identification approach cannot identify multiple components of a mixture. In most cases, e.g., in protein spots from 2D gels, the major component of a protein mixture is identified using MALDI-TOF from one single spot. Highly homologous proteins are sometimes difficult to distinguish with current software (84). Another problem is that it is not possible to directly identify potential biomarkers. Signals such as acute-phase responses or artefacts cannot be filtered out and may



be identified as biomarkers. To handle this limitation, tandem MS techniques and/or cumbersome purification is required.

Despite this limitations, MALDI-TOF-MS and ESI-MS/MS, combined with LC, are widely used for clinical proteomics and have contributed significantly to the human proteome organization-Plasma Proteome Project (HUPO-PPP) which has aims of creating a comprehensive database of all plasma and serum constituents and characterising the sources of variation within individuals over time. The collaborative effort of the HUPO initiative has analyzed standardized human plasma samples using various proteomic platforms. Through this effort, 3020 proteins have been qualitatively identified, each identification based on a minimum of two high-scoring MS/MS spectra. A critical step subsequent to protein identification is functional annotation. A subset of proteins from this project was annotated for relevance to cardiovascular disease. Most of the proteins in the vascular and coagulation system and markers of inflammation have been shown to localize in plasma, whereas the majority of other groups such as signaling, growth and differentiation, cytoskeleton, transcription factors or channels and receptors hosted a larger number of novel plasma components. Knowledge of the role of these plasma constituents on the cardiovascular system can provide insights into their roles in the plasma (85). The study of Donahue et al. led to the discovery of proteins related to coronary artery disease using large-scale proteomic analysis of pooled plasma. They analyzed 53 males with angiographic coronary artery disease and 53 control subjects without coronary disease from the Duke Databank for Cardiovascular Disease. Major plasma protein abnormalities were excluded. Plasma samples from each group were pooled to identify low-abundance proteins. After removal of albumin and immunoglobulins, and enrichment of smaller proteins (<20–40 kDa), samples were separated into 12,960 fractions by one cation exchange and two reversed-phase chromatography steps. Proteins were analyzed using LC-ESI-MS-MS. They could identify 731 plasma proteins or fragments thereof. Of these proteins, 95 were different between cases and controls. These represent broad categories of proteins involved with natural defence, inflammation, growth, and coagulation (86).

**Surface enhanced laser desorption/ionization (SELDI)** SELDI-TOF-MS, first described by Hutchens and Yip, is a hybrid technology combining chip-based solid phase chromatography with TOF-MS. In brief, incubating samples with chips whose surfaces are coated with a protein-fractionating resin allows certain proteins to become attached to the chips (87). After washing the unbound components away, an energy absorbing matrix is layered over the chips. Spectra are acquired using laser ionization and TOF separation MS. This allows protein separation and MS analysis to be performed using the same analysis system. Also, due to the chip format, high-throughput of up to 80 samples/day using current systems, low

sample volume (1  $\mu$ L; 25–50 cells, peptides in the fmol range) and relatively short analysis time can be achieved. Chips contain a chromatographic surface that can enrich protein subsets according to predefined conditions on a given surface; combining different surfaces and conditions facilitates the comprehensive analysis of complex protein mixtures.

A great number of studies have been published in recent years with currently close to 600 Medline entries for SELDI. Many of these studies have been directed towards the search for new biomarkers. However, studies on protein-protein and protein-DNA interactions, transcription factors and protein phosphorylation have also been performed. The possibility of binding specific antibodies to the chip also resulted in reports of the identification of isoforms and modifications of proteins such as troponin I and transthyretin. The commercial availability of this system and the advantages mentioned above regarding fast sample analysis are the main reasons why such a great number of studies attempting to translate proteome analysis in human specimens to clinical diagnostics have been performed using the SELDI platform. Consequently, many of the general problems of the proteomic approach regarding pre-analytical and analytical issues, as well as data interpretation and validation of results have been raised and discussed in great detail with respect to SELDI. Also, there are a number of methodological criticisms specific to the SELDI platform that deserve close attention and must be considered when comparing and evaluate systems designed for clinical applications of proteome analysis. These methodological criticisms, along with the basic principles and limitations of the method, have been reviewed extensively by Poon (88). The quality of the mass spectrometer component of available systems is under debate, especially its effect on resolution and sensitivity. Another point is the fact that proteins cannot be identified directly in the SELDI platform, instead requiring mass patterns for diagnosis. There is consensus, and it has been shown in a number of studies, that utmost efforts must be made to standardize and wherever possible automate the entire analysis to obtain acceptable precision and reproducibility. Bearing these limitations in mind, and considering the general precautions set out in this document, it should be noted that at present the majority of studies describing the application of protein marker patterns in clinical studies have been obtained with this technique. Advances in developing other more robust and precise methods that can be used for high-throughput analysis of small samples may change this situation.

**Array-based proteomics** High-density DNA microarray technology has played a key role in the analysis of the whole genome and gene expression patterns. Protein arrays are emerging to follow DNA microarrays as a potential screening tool for identifying protein-ligand interactions, and have great potential as a research and diagnostic tool for parallel processing of complex samples.

For array-based proteomics, molecular probes designed to capture specific proteins at specific sites are deposited on a solid support. To date, a large repertoire of solid supports based on glass-, plastic-, PVDF- or silicon-slides, often with chemically modified surfaces, are available. Molecular probes can be monoclonal or polyclonal antibodies, robust affinity proteins based on the structure of protein A (antibodies), highly thermostable members of large combinatorial libraries, mimicking natural ankyrin repeat proteins in *E. coli* cultures (ankyrin repeat proteins) or short lengths of single-stranded DNA or RNA molecules (aptamers). Arrays are probed with cell culture supernatants, cell lysates or serum. Depending on the molecular probes used, proteins or antibodies in the sample are bound by the planar array. The bound molecules are detected by a secondary antibody marked with a fluorescent dye, or directly if the sample has been fluorescently labeled. The incubated chips can be read by a variety of scanners based on non-confocal, confocal and planar wave guide technology (89).

The determination of autoantibody profiles with protein biochips promises to be a valuable part of future clinical diagnostics. Autoantibodies have already proven their usefulness in routine clinical diagnosis as surrogate and non-surrogate biomarkers. Some have been shown to be useful for disease monitoring and can be detected years before the onset of a disease (90). The systematic identification of potential disease-specific antigens opens the possibility for new diagnostic and therapeutic tools and, therefore, potential for economic gain. Screening sera or plasma from patients with protein arrays would not only allow the identification of potentially new antigens, but also enable the diagnosis and subtyping of autoimmune diseases based on the presence of specific auto-antibodies. This had lead to the profiling of the antibody repertoire of patients with various diseases. In a series of pioneering studies a number of putative novel autoantigens have been identified from one of the largest collections of recombinant human protein expression clones (Uniclone® technology, www.protagen.de) using pools of patient sera (91–93). The UNiclone® collection consists of ~ 11,000 different, sequence-characterized human recombinant proteins. In a different approach, a protein array consisting of 196 structurally diverse biomolecules representing major autoantigens was probed using serum from patients with various autoimmune diseases including systemic lupus erythematosus, Sjögren syndrome and rheumatoid arthritis. There were distinct autoantibody patterns for the different autoimmune diseases suggesting their utility for diagnosis (94). A new technique, termed layered peptide array, can serve as a screening tool to detect antibodies in a highly multiplexed format. The prototype was capable of producing ~ 5000 measurements/experiment. For Sjögren syndrome, this platform exhibited both a high-sensitivity (100%) and high-specificity (94%) for correctly identifying Sjögren syndrome antigen B antigen-positive samples from

patients with Sjögren syndrome (94, 95). Apart from tumor marker or autoantibody identification, cytokine networks in inflammation or transplant rejection may be detected by planar or bead-based protein arrays.

Following the development of disease-specific or disease group specific planar protein arrays, clinical laboratories can easily adopt such protein biochips because they are compatible with established and affordable DNA microarray scanners. Furthermore, new biochip platforms allow handling of multiple samples in a microtiter plate like format.

In addition to diagnostics, planar arrays can also be applied for the development of therapeutic antibodies. In this context, a protein biochip tool has been developed to speed up antibody research and to reduce the risk of failure based on its UNiclone® collection. These commercially available biochips (UNichip® AV-400 and UNichip® AV-VAR, www.protagen.de) are used to determine sensitivity, epitope specificity and the level of cross-reactivity of these antibodies. An added value is "faster selection of the best" antibodies prior to expensive animal experiments or clinical trials. Protein microarrays belonging to the UNichip® series will be applied to further interaction studies such as screening for protein kinase substrates, protein-protein interaction as well as protein-small molecule interaction.

#### **Bead-based immunoassays for protein analysis**

Apart from planar microarrays, bead-based systems provide an alternative when the number of parameters to be determined in parallel is rather low. In bead-based assay systems, latex microspheres are used as the solid phase and allow rapid binding kinetics and facilitate the separation step (1). Microsphere based assays have become an attractive alternative to the popular microtiter plate based, enzyme linked immunosorbent assay (ELISA). Sensitivity, reliability, and accuracy of microsphere based arrays are similar to those observed with well-established ELISA procedures. In fact, since 10,000 beads or more are measured for each analyte, and since each bead can be regarded as a single immunoassay, the precision of the test results is usually very favorable. Bead-based assay systems also have several advantages over planar microarrays (96). Probe molecules can be conjugated with millions of microspheres with high reproducibility. The composition of the panel of test parameters can be defined by the user by simply adding or removing beads with different probes; with a planar array the probe molecules are fixed. In addition, binding of the molecules in patient samples can be accelerated by mixing the probes, which is not always possible using planar protein arrays.

In contrast to planar arrays, probe molecules cannot be coded by their position on the protein chip. In steady, they have to be coded by variations in the microspheres they are attached. One example would be by variations in color or size. Most of the available assay systems use color coded microspheres. The beads are filled with one or more fluorescent dyes and can be measured with a flow cytometer. The

amount of captured target protein is quantified using a reporter system. Although this methodology is well-suited for single-analyte analysis, it is more desirable to rapidly and simultaneously quantify multiple analytes using a relatively small sample size. Sample size becomes a critical factor in the evaluation of multiple analytes. This technique enables multiplexed immunoassays with the use of multiple microspheres that can be discerned by different fluorescent labels as solid supports that are coated with specific antigens or antibodies. A set of a hundred different color coded beads is commercially available for multiplexed ligand-binding assays ([www.luminexcorp.com](http://www.luminexcorp.com), [www.bdfacs.com](http://www.bdfacs.com), [www.illumina.com](http://www.illumina.com)). For example, Luminex Corp offers microspheres coded by two dyes at ten different concentrations. Thus, up to 100 different sets with each matched to a different probe molecule can be used (97, 98). Such systems have been used to determine the concentration of cytokines (97, 99), screening for cystic fibrosis (97, 100), hepatitis B seroconversion and human immunodeficiency virus (97, 101, 102), thyroid hormones (97, 103), kinase testing (97, 104), allergy testing (97, 105), single nucleotide polymorphisms (97, 106), infectious disease diagnosis (97, 107), and detection of biological warfare agents (97, 108) or antibodies in serum or cell culture supernatant (109, 110).

Another possibility for distinguishing different beads is to generate microspheres of different sizes. These can be differentiated by light scatter. Combining both forms of beads, one could scale up to many more parameters in one test (97). Multiplexed analysis of protein-protein interactions (111), or the simultaneous characterization of the binding of auto-antibodies to multiple epitopes (112) are promising applications of flow cytometric bead-based assays.

In conclusion, bead-based immunoassays are widely used in the routine laboratory setting for the detection of cytokines or autoantibodies in serum (109). Due to the high-specificity and high-sensitivity, bead-based microarrays are currently a primary assay platform in clinical proteomics. They are easy to handle and the analyzers are affordable for most laboratories.

## Applications of proteomics and peptidomics of body fluids for the clinical laboratory

### Preparation of clinical samples for fluidic proteomics

Considering the multiple challenges arising from the complexity of disease processes, the heterogeneity and variability of human clinical samples, and the very low concentration of potential biomarkers in plasma, it is still questionable whether a direct proteomics approach, such as MS/MS, will reliably detect disease biomarkers in unfractionated plasma or serum in a routine clinical setting. Therefore, it is necessary to distinguish clinical proteomic applications from clinical proteome studies in medical research directed at understanding rather than diagnosing disease. This distinction is pivotal in any discussion on

sample sources, sampling, sample quality, and the entire pre-analytical and analytical process in proteomic studies.

With a clear understanding of the task at hand, it is mandatory to establish SOPs adapted to the specific requirements prior to collecting samples. SOPs must cover the entire process comprising collection, handling, preparation and storage, analysis of samples, including isolated cells and plasma, liquor, bronchoalveolar lavage (BAL) and urine (113, 114). It has been proposed that collection of urine is less challenging as there is little proteolytic degradation for several hours and, therefore, no additives are required (115). However, this has yet to be substantiated. Using a standardized protocol employing magnetic bead separation and MALDI-TOF-MS, Fiedler et al. detected 427 different mass signals in the urine of healthy donors and found acceptable within- and between-day imprecision (116). However, they worked with highly abundant proteins and this statement may not be true for less abundant or rare proteins. Well designed biomaterial repositories using on carefully planned collection and storage of clinical samples that are annotated with high-quality, disease-related clinical data (case history, diagnostic phenotype, treatment scheme, etc.), and maintained in a standardized format that can be used to compare results across studies will be a prerequisite for biomarker validation. It must be emphasized that problems with preanalytical standardization are prone to arise in the time period between collection of the biomaterial and its arrival at the laboratory. Quality management concepts such as the international accreditation norm EN/ISO 15,189 for medical laboratories proposes to place the preanalytical phase under the responsibility of the clinical laboratory. With compliance from clinical units, such regulations may improve the preanalytical phase. In their recent reviews, Luque-Garcia (4) and Paik (3) elaborate on the issues of clinical proteomics with emphasis on fluid proteomics and the preparation of clinical specimens and technological aspects of proteomic profiling.

**Plasma/serum** In laboratory diagnostics, human blood is the most frequently analyzed sample as serum, plasma or separated blood cells. Since blood is in contact with every organ and, therefore, contains proteome subsets of other tissues, it is the most complex human derived proteome. The number of serum proteins is estimated to be 10,000.

In principle, proteomics shows great promise for the study of proteins in plasma and a number of proteomic databases have already been established. A challenging problem of plasma proteome analysis is the dynamic range of protein concentrations which can be as much as 10 orders of magnitude. For example, albumin is present in blood at millimolar ( $10^{-3}$ ) concentration while many cytokines, such as tumor necrosis factor (TNF), are physiologically active at concentrations between  $10^{-12}$  mol to  $10^{-9}$  mol (117). Furthermore, only 10 proteins constitute 95% of the entire mass of plasma proteins (118). Since biomarker

discovery means searching for low-abundance proteins, the efficiency of fractionation systems are essential to avoid interference from abundant proteins. Therefore, development of multi-dimensional fractionation methods is vital to overcome the effects of ion quenching which is responsible for insufficient diagnostic sensitivity in unfractionated material. Two major depletion methods are used; resin based and antibody based depletion (e.g., multiple affinity removal system MARS). Several different techniques for plasma protein fractionation have been employed including 2D liquid enrichment system (Gradiflow®), plasma fractionation using multichannel electrolyte (MCE) and microscale solution isoelectric focusing (IEF) ZOOM, and free flow electrophoresis (FFE) (119, 120). However, caution is needed so that removal or pre-fractionation steps will not disturb the quantitative composition of the original sample and may misleading results.

In conclusion, obtaining proteomic profiles of uncharacterized and unidentified molecules is certainly advantageous in comparison with standard immunoassay measurements since a protein fingerprint can be obtained rapidly from as little as 1  $\mu$ L of unfractionated patient serum. This small sample can be analyzed using MS approaches to rapidly generate a unique proteomic signature of the serum (121). A general limitation of this approach is that MS-based methods select for the most abundant peptide ions and peaks making it likely that markers present in minute amounts are left undetected. Current immunoassays are able to detect low abundant proteins at a sensitivity of up to  $10^{-14}$  or even  $10^{-17}$ .

**Urine** The composition of urine is highly variable. The variability of its composition is a major way in which the "milieu interieur" is kept more or less constant in the face of environmental and nutritional changes. General problems with urine in laboratory diagnostics also apply to clinical proteomics. These problems consist of: 1) difficulties in standardizing sample collection and handling, 2) volume changes related to concentration changes and salt composition which vary widely, and 3) frequent contamination, especially in 24 h urine. In disease, there is frequently proteinuria and bacteriuria even in apparently healthy individuals. On the other hand, urine protein diagnostics has used pattern approaches and qualitative analyses for a long time.

The filtered plasma "peptidome" is normally processed by the proximal renal tubule. The proximal tubule removes substantial but undefined amounts of peptides and proteins from the filtrate. In addition, proteins are shed from the urinary track wall and from the bladder epithelium. Numerous cellular and membrane proteins are found in normal urine. When the proximal tubular reabsorption process is ineffective, as in the renal Fanconi syndrome, large quantities of plasma peptides are found in urine. In order to use urine as a source of reliable peptide biomarkers in disease, one has to define how variables within urine itself ("endogenous variables"), such as salt

composition and pH, influence peptide recoveries. In addition, sample processing variables ("exogenous variables"), such as freeze-thaw cycles, affect results. A single freeze-thaw cycle can produce dramatic changes in the intensity of several urinary peptides. The "urinary peptidome" promises to be a resource at least as dynamic and informative as the "urinary proteome". However, urine, as a matrix, is one of the least desirable biological fluids for both peptidomic and proteomic work. There are three main problems: (a) as is the case for proteins, the range of peptide concentrations in urine spans several orders of magnitude, (b) we still cannot quantify most of the peptides, and (c) beyond their mass measurement, we do not know their structure. A significant difficulty is the presence of large quantities of uromodulin ("Tamm-Horsfall protein") in urine. Uromodulin is the single major protein in healthy urine. The problem is that uromodulin forms fibrils which in turn form sediment depending on salts and pH. This is important because uromodulin is known to bind several LMW proteins and plasma peptides that enter the tubular filtrate. The elucidation of disease-specific biomarkers in urine is complicated by significant changes in the urinary proteome during the day. These changes are likely due to exercise, variations in diet and circadian rhythms (122). Thus, the reproducibility of biological assays is reduced because of physiological changes, and not due to poor reproducibility of the analytical method. In addition, differences between first void and midstream samples have been noted (Mischak et al., personal communication), highlighting the importance of standardized protocols for the collection. Certainly, we do not always have the choice how to collect urine. When considering urine collection in babies non-invasive sampling is preferred compared to biopsy. Apparently CE-MS, due to the detection of large numbers of peptides, is less sensitive to these variations than other methods. This variability in protein profiles highlights the difficulties in establishing a "normal human urine proteome map". Using urine samples from healthy volunteers following acetone precipitation, Thongboonkerd et al. (123) defined the first human proteome map, consisting of 67 proteins and their isoforms that could be used as a reference. In a subsequent study by Oh et al. (124), pooled urine samples from 20 healthy volunteers were used to annotate 113 proteins using 2-DE peptide mass fingerprinting (PMF). Additional experiments that further expanded the knowledge of the normal urinary proteome have been reported and ~ 800 proteins have been identified in the urine proteome (125, 126). Mann and co-workers identified more than 1543 proteins in normal urine pooled from ten people using gel and LC Fourier transform (FT)-MS/MS and Orbitrap MS/MS (127).

A direct comparison of identical urine samples using SELDI with CE-MS by Neuhoff et al. (128) resulted in the identification of three potential biomarkers using SELDI, and 200 potential biomarkers using CE-MS analysis. The authors concluded that it is necessary to characterize any disease by using a panel of



well-defined biomarker proteins rather than a few ill-defined peaks. Mischak and co-workers (129–131) used CE coupled to MS, together with appropriate software solutions, to analyze urine and other body fluids to diagnose various kidney disorders based on well-defined protein patterns. With this approach each protein is defined by its mass and migration time, and the signal intensity serves as measure of its abundance (70, 132). The urine samples were analyzed individually and the data from the individual CE-MS runs were combined due to the high reproducibility of the method. This feature allows compilation of datasets and comparison of the different groups; for example patients with a specific kidney disease compared to patients with other types of kidney disease or healthy controls. This comparison allows evaluation of an array of biomarkers that differentiate healthy subjects from patients, as well as other markers that define the specific disease or clinical condition. The latter type of biomarkers is useful for differential diagnosis. CE-MS permits fast and reproducible analysis and differentiation of protein patterns based on dozens of protein markers. Panels of 20–50 protein markers enabled diagnosis of a specific (primary) kidney disease as well as discrimination between different kidney diseases with high-sensitivity and specificity such as IgA nephropathy, focal-segmental glomerulosclerosis, membranous glomerulonephritis, minimal change disease, and diabetic nephropathy (133, 134). In this context, urine proteome diagnostics may represent a diagnostic approach to kidney disease without significant morbidity when compared with invasive kidney biopsy. In a recent study Decramer et al. (135) used CE-MS based urinary proteome analysis to define specific biomarker patterns for different grades of ureteropelvic junction obstruction, a frequently encountered pathology in newborns. Of note, these patients did not have any sign of increased proteinuria. In a blinded prospective study, these patterns predicted, with 95% accuracy, the clinical outcome of these newborns 9 months in advance. This data clearly indicate the potential of urinary proteomics for diagnosis as well as prognosis of renal disease. Proteome analysis of urine has also revealed biomarkers for several non-renal diseases. As in the case of ureteropelvic junction obstruction, these diseases generally do not result in increased proteinuria. Not surprisingly, biomarkers for urothelial cancer have been detected in urine. While the first studies using SELDI technology analyzed few samples and reported different biomarkers for the same disease (136, 137), Theodorescu et al. (115) recently used CE-MS to assay more than 600 samples, including 180 samples examined in blinded fashion as a validation set. The biomarkers found by these investigators correctly classified all blinded urothelial cancer samples and normal controls. However, nine of 138 patients with various chronic kidney diseases or nephrolithiasis were incorrectly classified as having urothelial cancer. Kaiser et al. (138) found biomarkers for graft-versus-host disease (GvHD) following bone marrow transplantation

using CE-MS-based urine proteomics. GvHD leads to endothelial dysfunction, which may also alter kidney structure and/or function. This complication can affect filtration and urine production resulting in disease-specific proteins being excreted in the urine.

**Other body fluids** Body fluids such as CSF, bronchoalveolar lavage (BAL), lacrimal, salivary, pleural, pericardic, pancreatic and synovial fluid, ascitis, bile and semen also are interesting clinical samples for proteomic analysis. CSF samples are collected by lumbar puncture. CSF has to be centrifuged to remove cells (e.g., at  $250\times g$  for 10 min) and the supernatant has to be stored at least at  $-80^{\circ}\text{C}$  with or without the addition of protease inhibitors. Because of the low protein concentration in CSF, sample preparation usually includes protein enrichment techniques such as ultrafiltration. Columns with different molecular weight cut-offs can be used alone or in combination. For example, Centricon YM-50 columns (Millipore, Bedford, CA) with a nominal molecular weight cut-off of 50 kDa may be used with Ultrafree columns (Millipore) with a molecular weight cut-off of 5 kDa. Potential serum biomarkers for early stroke have been identified using SELDI-TOF analysis. Analysis of CSF by MS has led to the identification of potential biomarkers for Alzheimer's disease (139) and potential biomarkers for stroke have been confirmed in serum of a large cohort of patients (140).

Problems with BAL samples include inconsistent dilution of the proteins in lung, depending on the fraction recovered. In addition, reference values from healthy donors are difficult to establish as BAL is indicated for patients with lung diseases only, and BAL is rarely performed on healthy probands. Over the years, the European Respiratory Society (ERS) task force has drafted documents on methods for performing BAL (141). BAL sample preparation and processing needs to be standardized and should follow international guidelines to make the results reproducible and comparable [International Scientific Societies of Respiratory Diseases (ATS/ERS) (142)]. BAL is usually performed during bronchoscopy for a variety of indications such as diagnosis and follow-up (143). Aliquots (50–60 mL) of phosphate buffered saline (usually four or five) are instilled by a fiber optic bronchoscope and the fluid is recovered by gentle aspiration. Recovery varies with lung site or type of disease. The first sample is generally kept separate from the others because it contains more debris and bronchial contamination and is not used for proteome analysis. The other aliquots are filtered and centrifuged to separate cells from the fluid component. The supernatant can be frozen at  $-80^{\circ}\text{C}$  until analysis. Cell differential counts are performed using cytocentrifuge preparations. The phenotype of lymphocytes and macrophages, or other cells, can be analyzed by flow cytometry using mAb (144). The study of BAL by use of sophisticated techniques such as proteomics requires better standardization of the method to reduce variability. For example, for 2-DE analysis, BAL samples must be desalted and concentrated to

obtain a suitable protein content. Factors linked to sample variability need to be considered. These factors include the quantity of fluid instilled, the minimal accepted percentage recovery, the type of aspiration, and the choice of site in the lung. Reference parameters are chosen to express the results, for example as mg of total protein, or albumin.

Diagnostic analysis of saliva for oral as well as systemic diseases depends on the identification of biomolecules that reflect characteristic changes in their presence or absence and the composition or structure of saliva components found in healthy and disease conditions. Most of the biomarkers suitable as diagnostic aids comprise proteins and peptides. The usefulness of salivary proteins for diagnosis requires recognition of typical features, making saliva unique among body fluids. Salivary secretions reflect a degree of redundancy displayed by extensive polymorphisms for families among each of the major salivary proteins. The structural differences among these polymorphic isoforms range from distinct to subtle, which may in some cases not even affect the mass of different family members. Knowledge of the structure and function of salivary derived proteins/peptides has a critical impact on the timely and correct identification of biomarkers, whether they originate from exocrine or non-exocrine sources (145).

Histological and functional changes of the lacrimal gland might be reflected in proteomic patterns in tear fluids. For example, determination of disease biomarkers in tear fluid for Sjögren's syndrome could lead to a non-invasive diagnostic test based on proteomic patterns. In a study with 31 Sjögren syndrome patients and 57 control subjects, protein profiling in tear fluids was identified using SELDI-TOF-MS. Multiple protein changes were detected reproducibly in the primary Sjögren syndrome group, including 10 potential novel biomarkers. Seven of the biomarkers ( $m/z$  2094, 2743, 14,191, 14,702, 16,429, 17,453, 17,792) were down-regulated and three biomarkers ( $m/z$  3483, 4972, 10,860) were up-regulated in the primary Sjögren syndrome group, comparing to the protein profiles of control subjects. When the cut-off value of the Sjögren syndrome score was set  $<0.5$ , 87% sensitivity and 100% specificity was achieved. The positive predictive value for this sample set was 100%. These findings support the potential of proteomic pattern technology in tear fluids for primary Sjögren syndrome (146). However, one disadvantage is that this technique is unable to identify specific proteins.

### Proteomic analysis of PTMs

Recent information on PTMs makes it possible to interpret biological regulation with new insights. Various protein modifications fine tune the cellular functions of each protein. Understanding the relationship between PTMs and functional changes is another enormous task, not unlike the human genome project. Proteomics, combined with separation technology and MS, makes it possible to dissect and characterize the individual parts of PTMs, and provide a systemic

analysis. Systemic analysis of PTMs of various signaling pathways has been applied to illustrate the kinetics of modifications. A variety of chemical modifications have been observed in proteins and these modifications alone, or in various combinations, occur in a time and signal dependent manner. PTMs of proteins determine their tertiary and quaternary structures and regulate their activities and functions. Recent advances in proteomic methodology, including MS, make it possible to identify proteins in complexes very rapidly (147). While protein identification can be accomplished using sequencing or mapping only a few peptides, mapping of PTMs requires the complete coverage of peptides comprising a protein. Protein modifications probably do occur in-vivo in more than 90% of proteins. Furthermore, the samples are a heterogeneous mixture of modified and unmodified proteins that are present in different proportions. Current proteomic technology is useful for detecting only simple modifications in large amounts of modified samples, not for thorough mapping of all endogenous protein modifications. Since proteomic methodology has tremendous potential for understanding PTMs, many efforts are being advanced for enriching modified samples and specific detection of modifications. Major types of PTMs are phosphorylation, acetylation, glycosylation, methylation, farnesylation, lipidation, GPI-anchors, sumoylation and ubiquitination (148, 149).

### Bioinformatic approaches in fluid proteomics

Although there are examples where a single laboratory parameter allows good clinical support for the diagnosis of disease, for example, troponin in myocardial infarction, most proteomic studies indicated that a single biomarker may be not adequate for reliable diagnosis, staging or prognosis of disease. The question arises of how to combine multiple biomarkers to provide a diagnostic or predictive pattern. Although a definitive answer is probably still far in the future, a number of approaches have emerged.

Among the first algorithms to utilize the available information on multiple biomarkers were hierarchical decision tree classification methods, such as Classification And Regression Trees (CART) (150). Heuristic clustering is another approach (151, 152). Empirical observations showed that these approaches are not successful because incorrect predictions made by the classification algorithm increased with the complexity of the decision tree. In addition, the number of datasets used for training the decision tree was low, resulting in a lack of statistical significance beyond the second or third nodes of the tree.

Support vector machines (SVMs), see reference (153) for example, seem to be a promising way to overcome some of these limitations due to the theoretical principles upon which they are based. In a number of diverse applications, excellent empirical performance of SVMs has been reported. Although mixed results were obtained with blinded datasets,

these approaches provide superior cross validated predictive performance. When the number of variables was  $<20$ , and substantial differences between the datasets existed, reliable results have been obtained. However, in cases where subtle differences exist, the number of support vectors need to be increased. This results in over fitting of the classifiers, to the training set and thus in poor classification of blinded datasets (Mischak et al., unpublished data). This term is also referred to as “memorizing”, a term often employed in artificial intelligence research. The number of variables and dimensions have to be decreased in order to avoid memorizing effects.

An important facet of the use of biomarker combinations for making predictive diagnosis with a classification algorithm is to have a properly calibrated indication of the level of confidence in the predictions being made. A classification such as, “this serum sample has been drawn from an individual with type II diabetes”, should also have a numeric score denoting how likely or probable it is that the classification is correct, i.e., “with 90% confidence this serum sample has been drawn from an individual with type II diabetes”. Of course, 90% confidence is more reliable than a prediction with 50% confidence, particularly if there are only two alternatives to be considered. In the case of presence of disease vs. absence, 50% confidence indicates little more than random guessing. Such confidence levels, also referred to as probabilities, attached to a classification enable costs of misclassification to be assigned in an optimal manner. Incorrectly predicting the absence of a disease has more serious consequences than to incorrectly predicting its presence. SVMs provide encouraging classification performance for a range of difficult problems. However, they are devoid of any probabilistic semantics and, therefore are unable to provide levels of confidence attached to any classification. Thus, the clinician is left with no information as to how much the predictions should be trusted.

A general purpose and computationally efficient Gaussian process based classification method has recently been developed (154). This method has been successfully applied to the problem of correct prediction of BRCA1 and BRCA2 heterozygous genotypes (155). This method provides a means of inferring optimally weighted combinations and possible selection of biomarkers.

Independent of which of these approaches is utilized, two basic considerations apply: 1) the number of independent variables has to be kept to a minimum and should be below the number of samples investigated, and 2) such an approach is only valid if applied to a blinded validation set, and it should be mandatory to include a blinded dataset in any report on potential biomarkers.

In a recent report, Rho et al. (156) presented a systems biology framework, called the “integrative proteomics data analysis pipeline” (IPDAP), which generates mechanistic hypotheses from network models reconstructed by integrating diverse types of

proteomic data generated by MS-based proteomic analyses. This framework includes a series of computational and network analysis tools and helps to translate data to biological knowledge.

### Data analysis strategies

In pre-clinical research, MS-based proteomics have become an important component. It is obvious that the analysis of complex protein or peptide mixtures, derived from body tissues or body fluids requires sophisticated data acquisition, handling and processing. Proteomic data needs to be combined and merged with patient or clinical data and compiled and integrated with very complex datasets. Highly heterogeneous data architectures are generated that need to be managed adequately. Data handling, interpretation, validation, storage and dissemination is critical to ensure proper use. Therefore, it is crucial to develop formats, as well as minimal requirements, to ensure data quality.

Current genomic and proteomic analytical methods, while highly developed and powerful, easily generate gigabyte-sized datasets. Informatics has to manage this data with respect to retrieval of information in a reasonable time and with appropriate quality. Furthermore, the abundant genomic or proteomic information that accumulated from prior studies could almost never be used adequately for the initial planning or interpretation of new experiments or data from a given study because the integration of data from outside studies was challenging and tedious. Therefore, it is not surprising that in every proteomic experiment results are rediscovered (157). The use of external and remote resources in own research causes various problems in many cases, as the access may be difficult and integration of the data retrieved is tedious or dependent on hardware. Very often, the compilation of the external information has to be done manually, and depends on the capabilities of the individuals involved. Transparency and efforts to generate the tools to integrate previous data into a current dataset, and experimental planning, will avoid redundant and costly studies. As suggested by Mewes (personal communication, Workshop Clinical Proteomics, Martinsried 9/2006), it would be preferable to combine resources with interfaces, using any desirable programming language, and use these interfaces to convert generic data formats into standard formats, send standardized information over the web, and thereby disseminate standardized and structured information to any clients via web services. The Human Proteome Organisation Proteomics Standard Initiative (HUPO PSI) has begun with these standardizing efforts already for proteomic datasets and the ProDaC consortium ([www.fp6-prodac.eu](http://www.fp6-prodac.eu)) is developing practical software tools to produce datasets in standardized formats. When data generation/analysis and conversion into standardized file formats is finished, export from the local database system into a central data repository like PRIDE (<http://www.ebi.ac.uk/pride>) will be obligatory for public datasets as well as private datasets following the publication of

results in scientific journals. Such central repositories will be fundamental in order to avoid rediscovery of results that are already known.

In ongoing biomarker discovery or clinical studies, large cohorts of individuals (patients and controls) are required in order to detect protein patterns that consistently associate with a specific condition and are distinguishable from the large background of proteins that randomly fluctuate within the population tested. The risk of misinterpreting correlating protein patterns as biomarkers is high and should be minimized where possible. This is particularly the case for a variety of cancer, but also for other diseases such as type 2 diabetes or heart failure where useful and adequate diagnostic markers with high-specificity and sensitivity are lacking, but urgently needed. Therefore, the success of MS-based discovery of candidate biomarkers depends on the ability to properly handle statistical data and interpret results with use of decoy databases and a defined false discovery rate. The value of the results obtained and the conclusions that are drawn can be limited by huge proteomic datasets, the heterogeneity of patients, sample processing and data acquisition in multi-centered studies, heterogeneity of data formats, processing methods, software tools and databases involved in the translation of spectral data into information. Since the beginning, biomarker discovery has suffered considerably from inconsistent data acquisition, statistical handling and validation. Clinical proteomics requires establishment of SOP and guidelines for specimen, reduction of complexity, increasing sensitivity of detection, application of bioinformatical tools for distribution of proteomic data into public databases, as well as thorough data generation (158, 159) consistent data analysis (92), and distinct training and validation of datasets (160).

These standards include:

Standards and data formats for data acquisition (raw data), storage (mzXML/mzData) and exchange (PSI: <http://www.psiview.info>).

- Public data repositories (Peptide Atlas, PRIDE, GPMDDB, SwissProt/Uniprot)
- Standards for quality assessment (e.g., FDR, composite decoy databases, power analysis)
- Integration towards a "linear pipeline" (Aebersold, ProDaC)
- Integration of complex databases including biological information (systems oriented approach)

When it comes to interpretation of complex data from clinically oriented studies, it is important to define rules of conduct for linking clinical phenotypes with profiling results, and to allow correct statistical evaluation and interpretation of protein profiles. Standards for data analysis and quality management have been established, standards for data reporting [HUPO-PSI, Molecular Cell Proteomics (MCP) and other journal guidelines] have been established, and databases for proteomics results are being structured within large HUPO initiatives. For the proteomic community, search algorithms and a public repository for

peptide identification has been developed (PRIDE: <http://www.ebi.ac.uk/pride>).

Data need to be qualified with respect to specificity for a pathological process. As a basis of proper data analysis, clustering variables, statistics, reproducibility, standards and exchange formats and study design have to be considered carefully. Sample retrieval, storage and handling are very important parameters that need to be strictly defined. Furthermore, standards and protocols need to be established to ensure reproducibility of the data. Recently established HUPO standards and SOPs for MS based data acquisition and analysis are considered important step towards this goal. However, a current challenge that still remains is heterogeneity of the data heterogeneity that might negatively impact integration of proteomics data with other biomedical data.

It is important to allow transparency in the handling of clinically relevant data by simultaneously ensuring protection of data and intellectual property. This limitation can only be overcome by teamwork amongst clinicians, bioinformatics, medical informatics and proteomic scientists.

Grouped analysis of proteomics data is important for scientific purposes and for development of efficient therapies. It is essential to use this type of data to assess therapy-dependent progression of disease in individual patients in advance, i.e., at diagnosis, disease relapse or other time points of interest. This should lead to individualized treatment of patients in stratified patient groups, and should maximize therapeutic success and minimize adverse drug reactions (personalized/individualized medicine) (161). The challenge relates to our ability to reach the right conclusions for short-, mid- or long-term therapeutic approaches, using dynamic proteome patterns that are influenced by various disease states.

Therefore, clinical proteomics studies need to include diseased patients as well as healthy or otherwise defined individuals as a reference group. In addition, an unknown validation set of diseased and reference individuals, including patients with unrelated diseases will help prove the robustness of classification for unknown patient samples.

## Conclusions

Even though there are a large number of bottlenecks in fluid proteomics. These include lack of standardization for specimen processing, quantitation, and clear strategies for managing biomarkers following their identification. It is believed that the field holds great promise. Progress depends on the establishment of SOPs for the selection of patients and specimens, decreasing the complexity of samples to be analyzed, and the development and use of superior informatics tools for efficient data management. The advances in proteomic analysis have made important improvements in the areas of sample fractionation, and parameter analysis relating to the fields of affinity binding technologies, high-resolution liquid phase



separation methods, as well as for liquid phase technologies like MS/MS, high-resolution MS-MS and advances in chemometry and biometry for large-scale data analysis. These form the basis for high-throughput and high-content analysis required for clinical proteomics, and offer the possibility for introducing these research tools into diagnostic research to screen for risk factors, identify new disease-specific or stage-specific biomarkers, and to find novel markers for therapeutic drug monitoring or new therapeutic targets. Therefore, clinical proteomics has the potential to complement genomics, metabolomics, lipidomics, glycomics and transcriptomics, including splice variant analysis to gain a better understanding of disease processes. The integration of proteomics and fluid and cell based technologies will ultimately lead to a description of the molecular setup of normal and abnormal cellular and liquid systems within a relational knowledge system. This will also allow standardized evaluation of abnormal disease states.

These methods are currently mostly qualitative and should be regarded as exploratory approaches that are advancing scientific knowledge within clinical studies, rather than routine. In the future, this will gradually change and some of the methods as well as applications will be established as clinical routine assays in the clinical laboratory. It is generally accepted that a set of different proteins or peptides (biomarkers), rather than a single protein or peptide, will be more efficient in the diagnosis of disease. Individualized prediction of the course of disease in patients using characteristic discriminatory data patterns will permit individualized therapies, identification of new pharmaceutical targets, and establishment of a standardized framework of relevant molecular alterations in disease (162). The control of preanalytical aspects is most important for clinical proteomics. This is best met by a high-degree of standardization, including SOPs, and automated work stations for high-throughput sample preparation.

## Acknowledgements

Work by the authors cited in this review was supported by the EU-projects SSA Lipidomics (ELIFE) Proposal-Nr.: 013032, FP7 EU project Lipidomic Net Proposal-Nr.: 202272 and the Danubian Biobank Proposal-Nr 018822.

## References

- Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. *Nature* 2003;425:905.
- Liotta LA, Petricoin EF. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest* 2006;116:26–30.
- Paik YK, Kim H, Lee EY, Kwon MS, Cho SY. Overview and introduction to clinical proteomics. *Methods Mol Biol* 2008;428:1–31.
- Luque-Garcia JL, Neubert TA. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *J Chromatogr A* 2007;1153:259–76.
- Lippi G, Salvagno GL, Montagnana M, Guidi GC. Reliability of the thrombin-generation assay in frozen-thawed platelet-rich plasma. *Clin Chem* 2006;52:1827–8.
- Conrads TP, Hood BL, Veenstra TD. Sampling and analytical strategies for biomarker discovery using mass spectrometry. *Biotechniques* 2006;40:799–805.
- Ferguson RE, Hochstrasser DF. Impact of preanalytical variables on the analysis of biological fluids in proteomic studies. *Proteomics-Clin Appl* 2007;1:739–46.
- Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. *Clin Chem Lab Med* 2008;46:764–72.
- Lippi G, Salvagno GL, Montagnana M, Brocco G, Guidi GC. Influence of hemolysis on routine clinical chemistry testing. *Clin Chem Lab Med* 2006;44:311–6.
- Guder WG. Haemolysis as an influence and interference factor in clinical chemistry. *J Clin Chem Clin Biochem* 1986;24:125–6.
- Sonntag O. Haemolysis as an interference factor in clinical chemistry. *J Clin Chem Clin Biochem* 1986;24:127–39.
- Grafmeyer D, Bondon M, Manchon M, Levillain P. The influence of bilirubin, haemolysis and turbidity on 20 analytical tests performed on automatic analysers. Results of an interlaboratory study. *Eur J Clin Chem Clin Biochem* 1995;33:31–52.
- Jay DW, Provasek D. Characterization and mathematical correction of hemolysis interference in selected Hitachi 717 assays. *Clin Chem* 1993;39:1804–10.
- Kroll MH, Elin RJ. Interference with clinical laboratory analyses. *Clin Chem* 1994;40:1996–2005.
- Steen G, Vermeer HJ, Naus AJ, Goevaerts B, Agricola PT, Schoenmakers CH. Multicenter evaluation of the interference of hemoglobin, bilirubin and lipids on Synchron LX-20 assays. *Clin Chem Lab Med* 2006;44:413–9.
- Hsieh SY, Chen RK, Pan YH, Lee HL. Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. *Proteomics* 2006;6:3189–98.
- Munro NP, Cairns DA, Clarke P, Rogers M, Stanley AJ, Barrett JH, et al. Urinary biomarker profiling in transitional cell carcinoma. *Int J Cancer* 2006;119:2642–50.
- Statland BE, Winkel P, Bokelund H. Factors contributing to intra-individual variation of serum constituents. 1. Within-day variation of serum constituents in healthy subjects. *Clin Chem* 1973;19:1374–9.
- Schaub S, Wilkins J, Weiler T, Sangster K, Rush D, Nickerson P. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney Int* 2004;65:323–32.
- Petri AL, Hogdall CK, Christensen IJ, Simonsen AH, T'Jampens D, Hellmann M-L, et al. Sample handling for mass spectrometric proteomic investigations of human urine. *Proteomics-Clin Appl* 2009;2:1184–93.
- Lippi G, Salvagno GL, Brocco G, Guidi GC. Preanalytical variability in laboratory testing: influence of the blood drawing technique. *Clin Chem Lab Med* 2005;43:319–25.
- Lippi G, Salvagno GL, Montagnana M, Poli G, Guidi GC. Influence of the needle bore size on platelet count and routine coagulation testing. *Blood Coagul Fibrin* 2006;17:557–61.
- Statland BE, Winkel P, Bokelund H. Factors contributing to intra-individual variation of serum constituents. 2. Effects of exercise and diet on variation of serum constituents in healthy subjects. *Clin Chem* 1973;19:1380–3.
- Banks RE, Stanley AJ, Cairns DA, Barrett JH, Clarke P, Thompson D, Selby PJ. Influences of blood sample processing on low-molecular-weight proteome identified by

- surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem* 2005;51:1637–49.
25. Baumann S, Ceglarek U, Fiedler GM, Lembcke J, Leichtle A, Thiery J. Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem* 2005;51:973–80.
  26. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 2005;5:3262–77.
  27. Tammen H, Schulte I, Hess R, Menzel C, Kellmann M, Mohring T, et al. Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. *Proteomics* 2005;5:3414–22.
  28. Sahu A, Lambris JD. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol Rev* 2001;180:35–48.
  29. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006;116:271–84.
  30. Albrethsen J, Bogebo R, Olsen J, Raskov H, Gammeltoft S. Preanalytical and analytical variation of surface-enhanced laser desorption-ionization time-of-flight mass spectrometry of human serum. *Clin Chem Lab Med* 2006;44:1243–52.
  31. Lewczuk P, Beck G, Esselmann H, Bruckmoser R, Zimmermann R, Fiszler M, et al. Effect of sample collection tubes on cerebrospinal fluid concentrations of tau proteins and amyloid beta peptides. *Clin Chem* 2006;52:332–4.
  32. Drake SK, Bowen RA, Remaley AT, Hortin GL. Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. *Clin Chem* 2004;50:2398–401.
  33. Mei H, Hsieh Y, Nardo C, Xu X, Wang S, Ng K, et al. Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: application to drug discovery. *Rapid Commun Mass Spectrom* 2003;17:97–103.
  34. West-Nielsen M, Hogdall EV, Marchiori E, Hogdall CK, Schou C, Heegaard NH. Sample handling for mass spectrometric proteomic investigations of human sera. *Anal Chem* 2005;77:5114–23.
  35. Engwegen JY, Alberts M, Knol JC, Jimenez CR, Depla AC, Tuynman H, et al. Influence of variations in sample handling on SELDI-TOF MS serum protein profiles for colorectal cancer. *Proteomics-Clin Appl* 2008;2:936–45.
  36. Carrette O, Burkhard PR, Hughes S, Hochstrasser DF, Sanchez JC. Truncated cystatin C in cerebrospinal fluid: Technical [corrected] artefact or biological process? *Proteomics* 2005;5:3060–5.
  37. Irani DN, Anderson C, Gundry R, Cotter R, Moore S, Kerr DA, et al. Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis. *Ann Neurol* 2006;59:237–47.
  38. Hansson SF, Simonsen AH, Zetterberg H, Andersen O, Haghighi S, Fagerberg I, et al. Cystatin C in cerebrospinal fluid and multiple sclerosis. *Ann Neurol* 2007;62:193–6.
  39. Bons JA, de BD, van Dieijen-Visser MP, Wodzig WK. Standardization of calibration and quality control using surface enhanced laser desorption ionization-time of flight-mass spectrometry. *Clin Chim Acta* 2006;366:249–56.
  40. Petricoin EF III, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002;94:1576–8.
  41. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62:3609–14.
  42. Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin Chem* 2002;48:1835–43.
  43. Mischak H, Apweiler R, Banks RE, Conaway M, Coon J, Dominiczak A, et al. Clinical proteomics: a need to define the field and to begin to set adequate standards. *Proteomics-Clin Appl* 2006;1:148–56.
  44. Marshall J, Kupchak P, Zhu W, Yantha J, Vrees T, Furesz S, et al. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J Proteome Res* 2003;2:361–72.
  45. Shaw J, Rowlinson R, Nickson J, Stone T, Sweet A, Williams K, et al. Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics* 2003;3:1181–95.
  46. Nedelkov D. Population proteomics: addressing protein diversity in humans. *Expert Rev Proteomics* 2005;2:315–24.
  47. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 2005;51:102–12.
  48. Klose J, Kobalz U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 1995;16:1034–59.
  49. Cristea IM, Gaskell SJ, Whetton AD. Proteomics techniques and their application to hematology. *Blood* 2004;103:3624–34.
  50. Rees-Unwin KS, Morgan GJ, Davies FE. Proteomics and the haematologist. *Clin Lab Haematol* 2004;26:77–86.
  51. Hoving S, Voshol H, van OJ. Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis* 2000;21:2617–21.
  52. Berggren KN, Schulenberg B, Lopez MF, Steinberg TH, Bogdanova A, Smejkal G, et al. An improved formulation of SYPRO Ruby protein gel stain: comparison with the original formulation and with a ruthenium II tris (bathophenanthroline disulfonate) formulation. *Proteomics* 2002;2:486–98.
  53. Lanne B, Panfilov O. Protein staining influences the quality of mass spectra obtained by peptide mass fingerprinting after separation on 2-d gels. A comparison of staining with coomassie brilliant blue and sypro ruby. *J Proteome Res* 2005;4:175–9.
  54. Nishihara JC, Champion KM. Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* 2002;23:2203–15.
  55. Page MJ, Griffiths TA, Bleackley MR, MacGillivray RT. Proteomics: applications relevant to transfusion medicine. *Transfus Med Rev* 2006;20:63–74.
  56. Thiele T, Steil L, Volker U, Greinacher A. Proteomics of blood-based therapeutics: a promising tool for quality assurance in transfusion medicine. *BioDrugs* 2007;21:179–93.
  57. Lilley KS, Razaq A, Dupree P. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol* 2002;6:46–50.
  58. O'Farrell PZ, Goodman HM, O'Farrell PH. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* 1977;12:1133–41.

59. Wittmann-Liebold B, Graack HR, Pohl T. Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. *Proteomics* 2006;6:4688–703.
60. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997;18:2071–7.
61. Alban A, David SO, Björkstén L, Andersson C, Sloge E, Lewis S, Currie I. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003;3:36–44.
62. Roe MR, Griffin TJ. Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* 2006;6:4678–87.
63. Stasyk T, Huber LA. Zooming in: fractionation strategies in proteomics. *Proteomics* 2004;4:3704–16.
64. Martosella J, Zolotarjova N, Liu H, Nicol G, Boyes BE. Reversed-phase high-performance liquid chromatographic prefractionation of immunodepleted human serum proteins to enhance mass spectrometry identification of lower-abundant proteins. *J Proteome Res* 2005;4:1522–37.
65. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17:676–82.
66. Washburn MP, Wolters D, Yates JR III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–7.
67. Mayya V, Rezaul K, Cong YS, Han D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Mol Cell Proteomics* 2005;4:214–23.
68. Tabb DL, MacCoss MJ, Wu CC, Anderson SD, Yates JR III. Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility. *Anal Chem* 2003;75:2470–7.
69. Johannesson N, Wetterhall M, Markides KE, Bergquist J. Monomer surface modifications for rapid peptide analysis by capillary electrophoresis and capillary electrochromatography coupled to electrospray ionization-mass spectrometry. *Electrophoresis* 2004;25:809–16.
70. Kolch W, Neuss C, Pelzing M, Mischak H. Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom Rev* 2005;24:959–77.
71. Hernandez-Borges J, Neuss C, Cifuentes A, Pelzing M. On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules. *Electrophoresis* 2004;25:2257–81.
72. Neuss C, Pelzing M, Macht M. A robust approach for the analysis of peptides in the low femtomole range by capillary electrophoresis-tandem mass spectrometry. *Electrophoresis* 2002;23:3149–59.
73. Theodorescu D, Fliser D, Wittke S, Mischak H, Krebs R, Walden M, et al. Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* 2005;26:2797–808.
74. Huang YF, Huang CC, Hu CC, Chang HT. Capillary electrophoresis-based separation techniques for the analysis of proteins. *Electrophoresis* 2006;27:3503–22.
75. Burkitt WI, Derrick PJ, Lafitte D, Bronstein I. Protein-ligand and protein-protein interactions studied by electrospray ionization and mass spectrometry. *Biochem Soc Trans* 2003;31:985–9.
76. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989;246:64–71.
77. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2299–301.
78. Salzano AM, Crescenzi M. Mass spectrometry for protein identification and the study of posttranslational modifications. *Ann Ist Super Sanita* 2005;41:443–50.
79. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004;20:1466–7.
80. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–67.
81. Yates JR III, Eng JK, McCormack AL, Schieltz D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 1995;67:1426–36.
82. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* 2003;3:1454–63.
83. Colinge J, Magnin J, Dessingy T, Giron M, Masselot A. Improved peptide charge state assignment. *Proteomics* 2003;3:1434–40.
84. Hortin GL. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin Chem* 2006;52:1223–37.
85. Berhane BT, Zong C, Liem DA, Huang A, Le S, Edmondson RD, et al. Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project pilot phase. *Proteomics* 2005;5:3520–30.
86. Donahue MP, Rose K, Hochstrasser D, Vonderscher J, Grass P, Chibout SD, et al. Discovery of proteins related to coronary artery disease using industrial-scale proteomics analysis of pooled plasma. *Am Heart J* 2006;152:478–85.
87. Simpkins F, Czechowicz JA, Liotta L, Kohn EC. SELDI-TOF mass spectrometry for cancer biomarker discovery and serum proteomic diagnostics. *Pharmacogenomics* 2005;6:647–53.
88. Poon TC. Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. *Expert Rev Proteomics* 2007;4:51–65.
89. Wingren C, Borrebaeck CA. Antibody microarrays: current status and key technological advances. *OMICS* 2006;10:411–27.
90. van Venrooij WJ, Zendman AJ, Pruijn GJ. Autoantibodies to citrullinated antigens in (early) rheumatoid arthritis [review]. *Autoimmun Rev* 2006;6:37–41.
91. Horn S, Lueking A, Murphy D, Staudt A, Gutjahr C, Schulte K, et al. Profiling humoral autoimmune repertoire of dilated cardiomyopathy (DCM) patients and development of a disease-associated protein chip. *Proteomics* 2006;6:605–13.
92. Stephan C, Reidegeld KA, Hamacher M, van HA, Marcus K, Taylor C, et al. Automated reprocessing pipeline for searching heterogeneous mass spectrometric data of the HUPO Brain Proteome Project pilot phase. *Proteomics* 2006;6:5015–29.
93. Lueking A, Huber O, Wirths C, Schulte K, Stieler KM, Blume-Peytavi U, et al. Profiling of alopecia areata autoantigens based on protein microarray technology. *Mol Cell Proteomics* 2005;4:1382–90.
94. Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ, et al. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 2002;8:295–301.
95. Gannot G, Tangrea MA, Gillespie JW, Erickson HS, Wallis BS, Leakan RA, et al. Layered peptide arrays: high-throughput antibody screening of clinical samples. *J Mol Diagn* 2005;7:427–36.



96. Wilson R, Cossins AR, Spiller DG. Encoded microcarriers for high-throughput multiplexed detection. *Angew Chem Int Ed Engl* 2006;45:6104–17.
97. Vignali DA. Multiplexed particle-based flow cytometric assays. *J Immunol Methods* 2000;243:243–55.
98. Kellar KL, Iannone MA. Multiplexed microsphere-based flow cytometric assays. *Exp Hematol* 2002;30:1227–37.
99. Kellar KL, Douglass JP. Multiplexed microsphere-based flow cytometric immunoassays for human cytokines. *J Immunol Methods* 2003;279:277–85.
100. Edelmann L, Hashmi G, Song Y, Han Y, Kornreich R, Desnick RJ. Cystic fibrosis carrier screening: validation of a novel method using BeadChip technology. *Genet Med* 2004;6:431–8.
101. Lukacs Z, Dietrich A, Ganschow R, Kohlschutter A, Kruihof R. Simultaneous determination of HIV antibodies, hepatitis C antibodies, and hepatitis B antigens in dried blood spots—a feasibility study using a multi-analyte immunoassay. *Clin Chem Lab Med* 2005;43:141–5.
102. Bellisario R, Colinas RJ, Pass KA. Simultaneous measurement of antibodies to three HIV-1 antigens in newborn dried blood-spot specimens using a multiplexed microsphere-based immunoassay. *Early Hum Dev* 2001;64:21–5.
103. Bellisario R, Colinas RJ, Pass KA. Simultaneous measurement of thyroxine and thyrotropin from newborn dried blood-spot specimens using a multiplexed fluorescent microsphere immunoassay. *Clin Chem* 2000;46:1422–4.
104. Luo Y. Selectivity assessment of kinase inhibitors: strategies and challenges. *Curr Opin Mol Ther* 2005;7:251–5.
105. Whitehead GS, Walker JK, Berman KG, Foster WM, Schwartz DA. Allergen-induced airway disease is mouse strain dependent. *Am J Physiol Lung Cell Mol Physiol* 2003;285:L32–42.
106. Hurley JD, Engle LJ, Davis JT, Welsh AM, Landers JE. A simple, bead-based approach for multi-SNP molecular haplotyping. *Nucleic Acids Res* 2004;32:e186.
107. Yan X, Zhong W, Tang A, Schielke EG, Hang W, Nolan JP. Multiplexed flow cytometric immunoassay for influenza virus detection and differentiation. *Anal Chem* 2005;77:7673–8.
108. McBride MT, Gammon S, Pitesky M, O'Brien TW, Smith T, Aldrich J, et al. Multiplexed liquid arrays for simultaneous detection of simulants of biological warfare agents. *Anal Chem* 2003;75:1924–30.
109. Fritzier MJ, Fritzier ML. The emergence of multiplexed technologies as diagnostic platforms in systemic autoimmune diseases. *Curr Med Chem* 2006;13:2503–12.
110. Stoll D, Templin MF, Bachmann J, Joos TO. Protein microarrays: applications and future challenges. *Curr Opin Drug Discov Devel* 2005;8:239–52.
111. Lund-Johansen F, Davis K, Bishop J, de Waal MR. Flow cytometric analysis of immunoprecipitates: high-throughput analysis of protein phosphorylation and protein-protein interactions 1. *Cytometry* 2000;39:250–9.
112. Tomer A, Koziol J, McMillan R. Autoimmune thrombocytopenia: flow cytometric determination of platelet-associated autoantibodies against platelet-specific receptors 1. *J Thromb Haemost* 2005;3:74–8.
113. Glocker MO, Ringel B, Gotze L, Lorenz S, Wandschneider V, Fehring V, et al. Klinische Proteomforschung. In: Gabrielczyk S, editors. *Laborwelt*. Berlin: Verlag der BIOCUM AG, 2000:7–12.
114. Sinz A, Bantscheff M, Mikkat S, Ringel B, Drynda S, Kekow J, et al. Mass spectrometric proteome analyses of synovial fluids and plasmas from patients suffering from rheumatoid arthritis and comparison to reactive arthritis or osteoarthritis. *Electrophoresis* 2002;23:3445–56.
115. Theodorescu D, Wittke S, Ross MM, Walden M, Conway M, Just I, et al. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* 2006;7:230–40.
116. Fiedler GM, Baumann S, Leichtle A, Oltmann A, Kase J, Thiery J, et al. Standardized peptidome profiling of human urine by magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem* 2007;53:421–8.
117. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–67.
118. Righetti PG, Castagna A, Antonucci F, Piubelli C, Cecconi D, Camprostrini N, et al. Proteome analysis in the clinical chemistry laboratory: myth or reality? *Clin Chim Acta* 2005;357:123–39.
119. Lee HJ, Lee EY, Kwon MS, Paik YK. Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. *Curr Opin Chem Biol* 2006;10:42–9.
120. Qian WJ, Jacobs JM, Liu T, Camp DG, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Mol Cell Proteomics* 2006;5:1727–44.
121. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. 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* 2005;5:3226–45.
122. Fliser D, Novak J, Thongboonkerd V, Argiles A, Janowski V, Girolami MA, et al. Advances in urinary proteome analysis and biomarker discovery. *J Am Soc Nephrol* 2007;18:1057–71.
123. Thongboonkerd V, McLeish KR, Arthur JM, Klein JB. Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney Int* 2002;62:1461–9.
124. Oh J, Pyo JH, Jo EH, Hwang SI, Kang SC, Jung JH, et al. Establishment of a near-standard two-dimensional human urine proteomic map. *Proteomics* 2004;4:3485–97.
125. Pieper R, Gatlin CL, McGrath AM, Makusky AJ, Mondal M, Seonara M, et al. Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* 2004;4:1159–74.
126. Sun T, Ye F, Ding H, Chen K, Jiang H, Shen X. Protein tyrosine phosphatase 1B regulates TGF beta 1-induced Smad2 activation through PI3 kinase-dependent pathway. *Cytokine* 2006;35:88–94.
127. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* 2006;7:R80.
128. Neuhoff N, Kaiser T, Wittke S, Krebs R, Pitt A, Burchard A, et al. Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Commun Mass Spectrom* 2004;18:149–56.
129. Kaiser T, Hermann A, Kielstein JT, Wittke S, Bartel S, Krebs R, et al. Capillary electrophoresis coupled to mass spectrometry to establish polypeptide patterns in dialysis fluids. *J Chromatogr A* 2003;1013:157–71.
130. Mischak H, Kaiser T, Walden M, Hillmann M, Wittke S, Herrmann A, et al. Proteomic analysis for the assessment of diabetic renal damage in humans. *Clin Sci (Lond)* 2004;107:485–95.



131. Wittke S, Fliser D, Haubitz M, Bartel S, Krebs R, Hausadel F, et al. Determination of peptides and proteins in human urine with capillary electrophoresis-mass spectrometry, a suitable tool for the establishment of new diagnostic markers. *J Chromatogr A* 2003;1013:173–81.
132. Fliser D, Wittke S, Mischak H. Capillary electrophoresis coupled to mass spectrometry for clinical diagnostic purposes. *Electrophoresis* 2005;26:2708–16.
133. Haubitz M, Wittke S, Weissinger EM, Walden M, Rupprecht HD, Floege J, et al. Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int* 2005;67:2313–20.
134. Weissinger EM, Wittke S, Kaiser T, Haller H, Bartel S, Krebs R, et al. Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int* 2004;65:2426–34.
135. Decramer S, Wittke S, Mischak H, Zurbig P, Walden M, Bouissou F, et al. Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med* 2006;12:398–400.
136. Liu J, Li M. Finding cancer biomarkers from mass spectrometry data by decision lists. *J Comput Biol* 2005;12:971–9.
137. Vlahou A, Schellhammer PF, Mendrinou S, Patel K, Kondylis FI, Gong L, et al. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. *Am J Pathol* 2001;158:1491–502.
138. Kaiser T, Wittke S, Just I, Krebs R, Bartel S, Fliser D, et al. Capillary electrophoresis coupled to mass spectrometer for automated and robust polypeptide determination in body fluids for clinical use. *Electrophoresis* 2004;25:2044–55.
139. Carrette O, Demalte I, Scherl A, Yalkinoglu O, Corthals G, Burkhard P, et al. A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics* 2003;3:1486–94.
140. Allard L, Burkhard PR, Lescuyer P, Burgess JA, Walter N, Hochstrasser DF, et al. PARK7 and nucleoside diphosphate kinase A as plasma markers for the early diagnosis of stroke. *Clin Chem* 2005;51:2043–51.
141. Haslam PL, Baughman RP. Report of ERS Task Force: guidelines for measurement of acellular components and standardization of BAL. *Eur Respir J* 1999;14:245–8.
142. Magi B, Bargagli E, Bini L, Rottoli P. Proteome analysis of bronchoalveolar lavage in lung diseases. *Proteomics* 2006;6:6354–69.
143. Magi B, Bini L, Perari MG, Fossi A, Sanchez JC, Hochstrasser D, et al. Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: a two-dimensional electrophoretic study. *Electrophoresis* 2002;23:3434–44.
144. Wahlstrom J, Berlin M, Skold CM, Wigzell H, Eklund A, Grunewald J. Phenotypic analysis of lymphocytes and monocytes/macrophages in peripheral blood and bronchoalveolar lavage fluid from patients with pulmonary sarcoidosis. *Thorax* 1999;54:339–46.
145. Oppenheim FG, Salih E, Siqueira WL, Zhang W, Helmerhorst EJ. Salivary proteome and its genetic polymorphisms. *Ann N Y Acad Sci* 2007;1098:22–50.
146. Tomosugi N, Kitagawa K, Takahashi N, Sugai S, Ishikawa I. Diagnostic potential of tear proteomic patterns in Sjogren's syndrome. *J Proteome Res* 2005;4:820–5.
147. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–61.
148. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–61.
149. Seo J, Lee KJ. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J Biochem Mol Biol* 2004;37:35–44.
150. Breimann L, Friemann J, Olshen RA, Stone JC. Classification and regression trees. Pacific Grove, CA: Wadsworth & Brooks/Cole Adv. Books Software.
151. Appel R, Hochstrasser D, Roch C, Funk M, Muller AF, Pellegrini C. Automatic classification of two-dimensional gel electrophoresis pictures by heuristic clustering analysis: a step toward machine learning. *Electrophoresis* 1988;9:136–42.
152. Hernandez P, Gras R, Frey J, Appel RD. Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data. *Proteomics* 2003;3:870–8.
153. Burges CJC. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining* 1998;121–67.
154. Girolami M, Rogers S. Variational Bayesian multinomial probit regression with Gaussian process priors. *Neural Comput* 2006;18:1790–817.
155. Kote-Jarai Z, Matthews L, Osorio A, Shanley S, Giddings I, Moreews F, et al. Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. *Clin Cancer Res* 2006;12:3896–901.
156. Rho S, You S, Kim Y, Hwang D. From proteomics toward systems biology: integration of different types of proteomics data into network models. *BMB Rep* 2008;41:184–93.
157. Pan S, Zhang H, Rush J, Eng J, Zhang N, Patterson D, et al. High throughput proteome screening for biomarker detection. *Mol Cell Proteomics* 2005;4:182–90.
158. Orchard S, Hermjakob H, Apweiler R. The proteomics standards initiative. *Proteomics* 2003;3:1374–6.
159. Orchard S, Hermjakob H, Julian RK Jr, Runte K, Sherman D, Wojcik J, et al. Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* 2004;4:490–1.
160. Domon B, Aebersold R. Challenges and opportunities in proteomics data analysis. *Mol Cell Proteomics* 2006;5:1921–6.
161. Valet G. Cytomics as a new potential for drug discovery. *Drug Discov Today* 2006;11:785–91.
162. Valet G, Leary JF, Tarnok A. Cytomics—new technologies: towards a human cytochrome project. *Cytometry A* 2004;59:167–71.