The handle http://hdl.handle.net/1887/25784 holds various files of this Leiden University dissertation.

**Author:** Nicolardi, Simone  
**Title:** Development of ultrahigh resolution FTICR mass spectrometry methods for clinical proteomics  
**Issue Date:** 2014-05-20
General Discussion
It is estimated that the human proteome contains more than a million different species, whereas the (current) number of human genes is approximately 20,300 [1]. The much higher complexity of the proteome than that of the genome is a result of not only gene expression itself but also of genetic variations, alternatively spliced RNA transcripts and a wide variety of post-translational modifications (PTMs). Proteins are generally considered as the work horses of the human cell and as such are a valuable source of information about physiological conditions. For this reason, disease-specific variations in proteins levels can be promising markers for (early) diagnosis or prognosis of the disease. In this context, although enormous progress has been made in the analysis (i.e. identification) of proteins, further scientific advancements are needed to link protein expression with disease (state). The need for new technologies and methodologies for the comprehensive analysis of the human proteome has driven the progress in the field of MS-based proteomics. Nowadays, a wide variety of separation techniques in combination with different mass spectrometers is available and used to obtain a detailed analysis of biological samples in terms of protein identity and quantity. However, proteome analysis is still limited by many factors, most importantly the ten orders of magnitude dynamic range in protein concentrations and the lack of information about different proteoforms. Consequently, worldwide, scientists have been and are still working on the development of new methodologies and technologies to further improve and innovate MS-based proteomics.

MS-based protein profiling was one of the first approaches used to demonstrate specific changes in the expression of peptides and proteins in samples from diseased individuals and has been widely applied ever since. Within a given human population protein expression is highly variable and consequently disease-specific changes can only be identified with high specificity when a significant number of human samples is analysed. This can most easily be achieved using automated sample processing methods that allow high-throughput screening. Apart from the aspect of high-throughput, robotic platforms follow a robust and standardized workflow and provide precise, reproducible and accurate data that are required for clinical discovery studies. For routine analysis in the clinical laboratories similar characteristics on robustness and standardization are required. Thus, further development of robust sample processing procedures is pivotal for implementation of MS-based proteomics strategies for clinical applications.
Cancer biomarker discovery studies based on serum peptide and protein profiling have been criticized because most of these discoveries could not be translated into a diagnostic clinical assay [2,3]. One common explanation for this setback is the inherent lack of “depth” in profiling studies, i.e. only the first three or four orders of magnitude of the protein concentration range are mapped [4]. Nevertheless, it may very well be that a layer of highly abundant proteins can still provide diagnostic or prognostic clues on a certain disease, as has been shown by others and is shown in Chapter 7 of this thesis [5]. The failed translation of initial MS-based biomarker studies more likely relates to a lack of standardized sample collection or issues with validation, as robustness and reproducibility of the analytical platform are essential [5-8]. With this in mind, we performed our MALDI profiling research as follows. First, a stringent protocol was followed for both the collection and storage of the human serum samples needed for our studies. Second, we developed and implemented two different automated liquid handling platforms in our workflow either based on functionalized magnetic beads or on cartridges packed with SPE material similar to that of the beads (Chapter 1). These platforms allowed the fractionation of serum peptides and proteins in a very reproducible way, thereby reducing the analytical variation generated in following processing steps. Both platforms provided similar MS results. It was concluded that the use of cartridges increased flexibility in sample preparation, since the cartridges can be packed with virtually any desired SPE material. In both cases, hundreds of samples can be processed in a few hours allowing the analysis of a large cohort of clinical samples in a few days.

The performance of the mass analysers that were used for the acquisition of the peptide and protein profiles had a clear effect on the results of the profiling studies. While MALDI-TOF-MS has been the most commonly used platform for peptide and protein profiling, MALDI-FTICR-MS has been rarely used for this purpose. Since this latter technology clearly offers better MS performance, we evaluated the applicability of this platform for profiling studies. To this end, a small cohort of human serum samples was analysed using both TOF and FTICR mass spectrometers and the results were compared. As shown in Chapter 2, part-per-million (ppm) mass measurement accuracy and precision were obtained by MALDI-TOF-MS after a baseline correction and internal calibration of the spectra generated from replicate measurements of 96 different serum samples. Measurement of the same set of samples using a 15 Tesla MALDI-FTICR mass...
spectrometer resulted in even lower ppm mass accuracy and precision. The sub-ppm mass measurement errors and the improved precision of the MALDI-FTICR-MS measurements were shown to lead to a more reliable identification of the peptides and simplified the analysis making the spectral alignment more accurate and robust. This improvement would appear to be a valuable characteristic for valid comparisons of profiles in clinical studies.

Despite the high standardization of used the sample preparation procedures and MS measurements, profiles with poor spectral quality can occasionally still be obtained because of several factors. For example, a “failure” of either the extraction procedure or the MALDI spotting can lead to profiles with an inadequate number of peaks, while the presence of contaminants will reduce the sensitivity of the platform with regard to measuring (lower abundant) peptides. The measurements with isotopic resolution in both MALDI-TOF and –FTICR MS experiments allowed the development of a quality control method for the selection of the best spectrum within the replicate measurements of one sample and for the removal of low quality spectra from further statistical analysis (Chapter 3). This method was based on the comparison of the observed isotopic distributions with the estimated polyaveragine distributions and allowed the semi-automated evaluation of a large cohort of samples, typical for clinical studies.

In Chapter 4, we describe the application of ESI-FTICR-MS/MS CID- and ETD-experiments for the identification of human serum peptides that are detected in MALDI-FTICR profiles. The ultrahigh resolving power of the FTICR measurements allowed the confident assignment of many fragment ions resulting in high sequence coverage while the sub-ppm mass measurements errors and precision made the identifications more reliable. CID and ETD results were shown to be complementary in terms of sequence coverage. The high mass measurement precision allowed a window of overlap between species observed in MALDI profiles and those determined in ESI spectra lower than 1 ppm.

Each single MALDI-MS profile obtained from a human serum sample is characterized by hundreds of features (i.e. peaks). The number of detected features depends on the sensitivity and the mass resolving power of the mass spectrometer. In reflectron TOF and FTICR mass analyzers, these characteristics significantly decrease with increasing m/z-values. Using a state-of-the-art, high-end reflectron MALDI-TOF-MS
peptides could be isotopically resolved up to \( m/z \)-value 4,500, while using a 15T MALDI-FTICR mass spectrometer it was possible to extend this range up to \( m/z \)-value 15,300 thus allowing the detection of a higher number of peaks. Moreover, MALDI-FTICR experiments clearly demonstrated MALDI-TOF overlapping peptides (Chapters 1-6). The development of an ultrahigh resolution MALDI-FTICR profiling method in the \( m/z \)-range from 6,000 to 15,300 allowed a more detailed investigation of endogenous peptides and small proteins detected in this \( m/z \)- range. The MALDI-FTICR-MS analysis of human serum samples in this mass range resulted in the detection of several apolipoproteins from the “C” family including glycosylated apolipoprotein-CIII isoforms (apoCIII’s), as is presented in Chapter 5. These isoforms contain different glycosylation on the Threonine 74 and have different ionization efficiency in MALDI, resulting in a lower signal intensity of the forms with one or two sialic acids. The evaluation of the apoCIII’s distribution in 96 different human serum samples showed a high biological variability and proved to be a powerful analytical tool. In view of the important role of apoCIII’s in various disease pathways, future studies will be performed using the developed MALDI-FTICR profiling platform to investigate a possible correlation of apoCIII’s with diseases in clinical cohorts. The identification of apolipoproteins and the other profiled peptides was performed by comparison with previously identified peaks in MALDI-TOF profiles and by MS/MS experiments. MALDI-TOF MS allows the identification of the profiled peptides using in source and post source decay (ISD and PSD) processes, collision-induced dissociation (CID) and their combination (e.g. LIFT, Bruker Daltonics). However, the identification of large polypeptides is limited by the lack of resolving power while the fragmentation of low intensity precursor ions usually leads to poor MS/MS spectra. Moreover, in MALDI-TOF experiments peptides are inherently singly charged, which by itself is a limiting factor for fragmentation studies. For these reasons, complementary ESI-MS/MS experiments of the same kind of fractionated human serum peptides were performed. In general, multiply charged species are less stable than singly charged ions and their fragmentation leads to a higher sequence coverage. Ion trap MS has been successfully employed to identify peptides that were previously separated by liquid chromatography (LC) [9]. Direct infusion ESI-MS/MS experiments can be performed, provided the complexity of the sample is sufficiently low. While LC-IT-MS/MS experiments allow the
identification of thousands of peptides per run, direct infusion experiments can be used to optimize the fragmentation parameters for those peptides that were not identified.

As reported in Chapter 6, the improved mass accuracy and precision of the ultrahigh resolution MALDI-FTICR-MS measurements allowed the identification of six new intact fucosylated apoCIII’s in the \( m/z \)-range from 6,000 to 15,300. These new glycoforms were less abundant than the normal apoCIII’s and showed a large variation in their relative abundance and frequency in the 96 studied serum samples. To corroborate these identifications, direct infusion ESI-FTICR-CID MS/MS experiments were used to elucidate the glycan moiety of one of the fucosylated forms while ion trap CID MS/MS experiments were performed to partially localize the modification site. It should be stressed that these apoCIII’s were previously observed in MALDI-TOF spectra, however not identified. In our study, the less abundant fucosylated forms could be detected as a result of the improved sensitivity and higher resolving power of the 15T MALDI-FTICR-MS.

The FTICR-MS system used to perform the experiments needed for our studies offered a large versatility of the mass measurements. Several parameters were optimized to improve the transmission of the ions from the ion source to the ICR cell and their detection. This resulted in different MS methods with enhanced sensitivity for specific \( m/z \)-ranges. Two of these methods were used to profile serum peptides and proteins in the \( m/z \)-ranges from 1,013 to 3,700 and from 3,500 to 10,000, respectively. These ultrahigh resolution MALDI-FTICR profiling methods were integrated with a fully automated, magnetic bead-based serum peptides and proteins fractionation protocol in a high-throughput and standardized profiling platform for clinical applications. In order to test the performance of this platform a cohort of serum samples from healthy volunteers and pancreatic cancer (PC) patients was analysed. In Chapter 7, we reported the result of this study. Serum peptides or small proteins could be measured with isotopic resolution and quantified with a higher accuracy and precision than any previously reported MALDI-TOF profiling study. PC discriminant proteomic signatures were identified after the statistical analysis of the profiles obtained from 273 serum samples. The specificities and sensitivities of these methods were in agreement with those previously reported by other authors. Many of the peptides observed in the MALDI-FTICR profiles have not yet been identified. Their characterization is important to understand their role in the biology of
human serum and the possible linkage with disease processes. The high mass accuracy of the 15T MALDI-FTICR measurements allowed the identification of 34 new peptides in the spectra.

In the field of MS-based proteomics, large efforts are made to identify proteins at very low concentration levels (secreted in serum), since these can be disease-specific early biomarkers. The profiling methods developed and described in this thesis primarily provide detection of high-abundance proteins and their proteolytic fragments, despite the improved dynamic range obtained using a 15T MALDI-FTICR-MS. Therefore, the development of fractionation techniques remains crucial for a more in-depth investigation of the serum proteome and the identification of a more specific peptide and protein signature which could be used as a valid disease marker. Different technologies and strategies are nowadays available to dig deep into the serum proteome. Unfortunately, these all lack the analytical robustness that is needed for clinical application. For example, the first 15-20 most abundant proteins can be removed by immunocapture using immobilized antibodies on chromatographic supports [10]. However, these techniques are not 100% efficient and their reproducibility has been criticized. Another example is the use of multidimensional separation-based bottom-up proteomics that on the one hand allows the identification and quantification of thousands of serum proteins but on the other hand has a very low-throughput thus limiting its application to only a small cohort of samples [9,11]. Another way of tackling the problem of the large dynamic concentration range of human serum proteins is by focusing the analysis only to specific protein(s) of interest [12]. To this end, targeted proteomics allows detection of a preselected group of proteins with high sensitivity, quantitative accuracy and reproducibility. Multiple-reaction monitoring (MRM) MS is highly sensitive and allows the accurate quantification of specific (often tryptic) peptides in a sample by using isotopically labelled internal peptide standards [13,14]. To further enhance sensitivity in an MRM-approach anti-peptide antibodies can be used to enrich specific peptides. This is commercialized by SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies), a method that is based on one or more specific anti-peptide antibodies to quantify a singly protein in a complex (biological) sample [15]. Recent attempts to improve the throughput and robustness of this technique were based on the use of robotic platforms for the magnetic bead-based immunocapture procedures and
MALDI-TOF MS for the measurements of the tryptic peptides. In this context, the ultrahigh resolution MALDI-FTICR MS methods developed for the studies and presented in this thesis can be used to pitch a SISCAPA platform for the accurate and precise analysis of small proteins and peptides. In addition, suitable antibodies can be used to enrich the isoforms of a specific protein allowing the evaluation of their distribution regarding a possible correlation with diseases [16].
REFERENCES


