

Cover Page



Universiteit Leiden



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

Development of ultrahigh resolution  
FTICR mass spectrometry methods  
for clinical proteomics

Simone Nicolardi

ISBN: 978-94-6182-435-6

The research described in this thesis was financially supported by the Center for Translational Molecular Medicine (CTMM).

Printing: OffPage, Amsterdam

Copyright © 2014 by S. Nicolardi. All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without prior permission of the author.

Development of ultrahigh resolution  
FTICR mass spectrometry methods  
for clinical proteomics

proefschrift

ter verkrijging van de graad van Doctor aan de Universiteit Leiden, op  
gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties te verdedigen op  
dinsdag 20 mei 2014, klokke 15.00 uur

door

**Simone Nicolardi**

geboren op 28 juni 1981, te Gagliano del Capo, Italië



## PROMOTIECOMMISSIE

**Promotores:** Prof.dr. A.M. Deelder

Prof.dr. R.A.E.M. Tollenaar

**Co-promotor:** Dr. Y.E.M. van der Burgt

**Overige Leden:** Prof.dr. G.A. Meijer  
*Dept. of Pathology, VU University Medical Center (VUMC),  
The Netherlands*

Prof.dr. R. Cramer  
*Dept. of Chemistry, University of Reading,  
United Kingdom*

Prof.dr. H.J. Tanke

Prof.dr. F. Koning

Prof.dr. A. van der Laarse

## TABLE OF CONTENTS

<b>Introduction</b>		<b>7</b>
<b>Chapter 1</b>	Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids	<b>29</b>
<b>Chapter 2</b>	Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry	<b>51</b>
<b>Chapter 3</b>	Quality control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling	<b>69</b>
<b>Chapter 4</b>	Identification of human serum peptides in Fourier transform ion cyclotron resonance precision profiles	<b>95</b>
<b>Chapter 5</b>	Mapping O-glycosylation of apolipoprotein C-III in MALDI-FT-ICR protein profiles	<b>109</b>
<b>Chapter 6</b>	Identification of new Apolipoprotein-CIII glycoforms with ultrahigh resolution MALDI-FTICR mass spectrometry of human sera	<b>131</b>
<b>Chapter 7</b>	Serum biomarker signatures of pancreatic cancer using MALDI-FTICR peptide and protein profiles	<b>161</b>
<b>General Discussion</b>		<b>191</b>
<b>Summary</b>		<b>205</b>
<b>Samenvatting</b>		<b>210</b>
<b>Acknowledgements</b>		<b>214</b>
<b>Curriculum vitae</b>		<b>216</b>
<b>List of publications</b>		<b>217</b>



# **Introduction**

A myriad of scientists have been fascinated from multiple perspectives by the complexity of blood. This biological fluid is known for a variety of functions such as transportation of oxygen and carbon dioxide through the lungs, transport and removal of nutrients, and heat transport and immunity. Blood collects products of the human metabolism throughout the entire body. The first studies on the blood circulation system can be dated back to 1260 when the Arabic scholar, mathematician and physician Ibn al-Nafis described the “minor circulation” of the human body [1]. Later, with the invention of the microscope, the Dutch naturalist Jan Swammerdam was the first one to observe red blood cells (1658) [2]. These cells have remained the only blood constituents observable with a microscope until 1842 when Alfred Donné, a French physician discovered platelets [3]. One year later, leukocytes were simultaneously described by Gabriel Andral, a French professor of medicine, and William Addison, an English country practitioner [4,5]. Nowadays we know that blood is a complex mixture of electrolytes and various biomolecules, such as lipids, carbohydrates, and proteins. Often, blood cells and fibrinogen are removed through centrifugation, yielding serum or plasma. The wealth of bioinformatic information, which can be derived from analysis of blood, in combination with the minimally invasive way to obtain blood by antecubital venipuncture have led to the availability of many (clinical) tests to routinely analyse specific serum or plasma components. The levels of biomolecules change according to the physiological condition of the body, although to a lesser extent than in the case of urine. One of the best known examples is the glucose concentration in blood, but there are many more species that can change in concentration or even identity over time. It is often hypothesized that such changes in the concentration of a biomolecule can be an indicator, or marker, of a specific

pathological state. To that end human serum or plasma samples have been studied extensively in relation to various diseases.

Like for other serum components, since a few decades abnormal proteins levels have been found to be associated with certain pathological conditions [6,7]. The variety of human serum proteins is vast, with over 10,500 gene products described (<http://www.plasmaproteomedatabase.org/> y2013). Protein concentrations in serum range over ten orders of magnitude [8]; these proteins all have different structures, amino acid compositions, molecular weights, chemical-physical properties and functionalities. Albumin is the most abundant serum protein (34-54 mg/ml in humans) and constitutes approximately 95% of the total serum protein content together with immunoglobulins, transferrin, macroglobulin, and apolipoproteins. This phenomenon severely hampers the analysis of the proteins present at much lower concentrations and although several analytical approaches can be used to decrease serum complexity, the investigation of these so-called low-abundant proteins remains challenging [8,9]. Serum complexity is furthermore increased by the fact that proteins can be cleaved by (endogenous) enzymes into smaller peptides or proteins. Specific enzymes are activated as a consequence of a change of the physiological condition cleaving specific peptide bonds and producing functional peptides or proteins. For example, angiotensinogen is cleaved by renin to form angiotensin I peptide, a ten amino acid long peptide that is further cleaved by angiotensin-converting enzyme to form angiotensin II peptide. This peptide is a hormone that increases blood pressure. Furthermore, many unspecific cleavages of proteins are observed in blood-related samples: peptides derived from such degradation are often referred to as “blood degradome”.

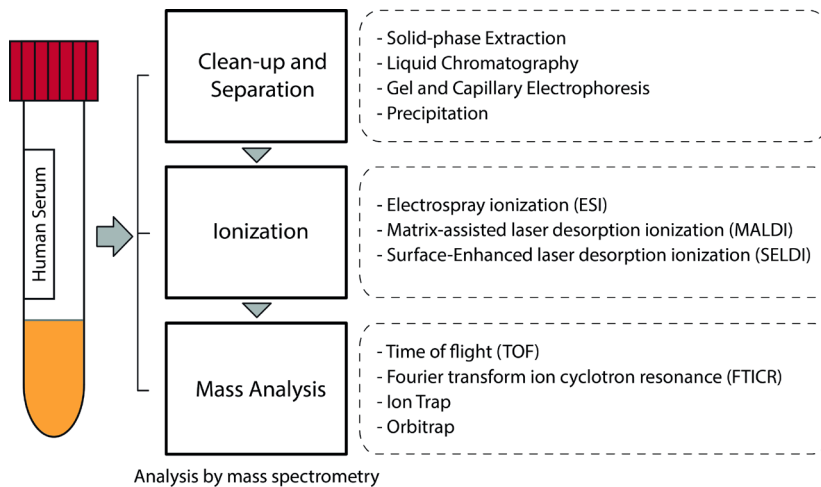
Methods for the determination of the elemental composition of proteins have been developed since the middle 1910s [10]. For this purpose, two-dimensional gel electrophoresis has been a widely applied technology to separate proteins [11]. It is commonly agreed that the identification of a protein from a gel by mass spectrometry (MS) through peptide sequencing was first reported by Matthias Mann [12]. However, several other scientists have pioneered in the field of peptide- and protein identifications by means of MS, like Sundqvist, Heerma, Roepstorff, Hunt and Yates [13-17]. It was only in the middle 1990s, at the same time that the Human Genome Project received great attention, that mass spectrometry for protein identification grew to its current size. It was clear that the identification and characterization of all human proteins would provide biologically relevant information complementary to genomic studies, however it would take a few more years before technology was in place. To make an analogy with the term genomics, this new discipline was called *proteomics* [18]. Consequently, new technologies were developed to provide innovative methods for the separation, the detection and the characterization of proteins, and MS has evolved from an early, laborious protein analysis tool [19,20] into a high-throughput highly automated method. Nowadays, MS is the method of choice in proteomics for the analysis of complex protein samples (*e.g.* human serum) [21]. However, its application requires proper and standardized sample preparation. There is an enormous variety of techniques that can be employed to clean up, separate or fractionate peptides and proteins from biological sources [22]. Reversed-phase (RP), ion exchange, size exclusion and affinity chromatography have been applied using different technologies including ZipTip® pipette tips, cartridges and magnetic beads for solid-phase extraction (SPE) and columns for liquid chromatography (LC). Only after the sample has been processed using one or more

of these techniques, it is ready for mass spectrometric analysis. MS-analysis requires an ionization source, a mass analyzer and a detector, usually integrated in a commercially available mass spectrometer or MS-system.

In a mass spectrometer, ions are generated in an ionization source and then transferred to the detector following a decrease of potential energy; during this path, ions can be trapped, focused, “cooled”, excited, accelerated, and fragmented. The technology behind all these processes is diverse and is used in different configuration to create mass spectrometers with different characteristics, resulting in different performance and capability. MS is sensitive, allows the detection of as little as a few hundred molecules and the identification of protein amino acid sequence including post-translational modifications (PTMs) such as glycosylation or phosphorylation. Despite this, the direct analysis of complex protein mixtures (*e.g.* human serum) by MS, with some exceptions (*e.g.* dried blood analysis, identification of bacteria), requires procedures that result in a reduced complexity of the sample (see Figure 1). The combination of these procedures with different ionization techniques and mass analysers allows the in-depth investigation of the human serum proteome.

The two most common ionization sources for biomolecules are based on matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI). It is fair to state that at least in part MALDI and ESI provide complementary information, and that the method of choice depends on the application. The latter ionization is generally used to analyse compounds present in solution and has the advantage of on-line coupling with chromatography systems. ESI of peptides and proteins produces a broad distribution of multiply-charged ions in the gas phase resulting in multiple peaks in the mass spectrum [23].





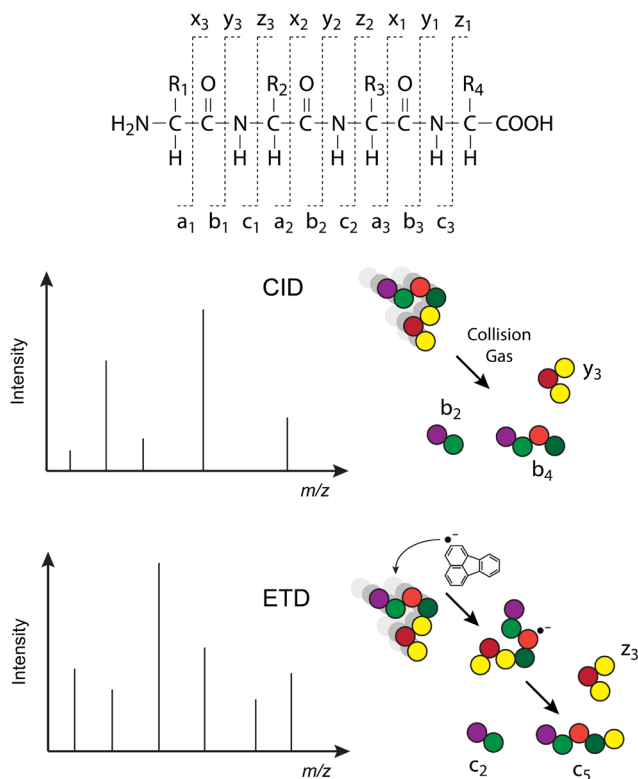
**Figure 1. General workflow for the MS analysis of human serum peptides and proteins.**

The charge state depends on both the composition of the liquid in which peptides and proteins are dissolved and their amino acid composition and structure. These multiply-charged ions are suitable for further MS identification experiments (*i.e.* MS/MS) that are based on the isolation and fragmentation of a particular ionized species (*i.e.* precursor ion). The preferred ions for MS/MS-analysis are (de)protonated species (*i.e.*  $[M+nH]^{n+}$  or  $[M-nH]^{n-}$ ), but also other precursor ions have been used for identification studies. The MS analysis of the fragment ions allows the identification of the primary sequence of the isolated peptide or protein. In so-called bottom-up proteomics studies, proteins are identified by digesting protein(s) into peptides with a proteolytic enzyme such as trypsin, followed by separation of the resulting peptides by RP-LC with continuous MS/MS “read-out” for each eluting peptide precursor. Peptide sequences are obtained from the corresponding MS/MS-spectrum with a database search tool, such as SEQUEST, Mascot, MS-GFDB, or Spectrum Mill.

Proteins identities (IDs) are then inferred from unique peptide sequence identifications. The utility of protein identification by database search depends upon the existence of a reference database that contains ideally all peptides of interest. This approach allows the identification of thousands of proteins in a single experiment [24]. However, it should be borne in mind that the presence of PTMs may affect the proteolytic efficiency of the enzyme, the recoveries from a clean-up procedure, retention time of the LC, and both the ESI the MS/MS efficiency. Thus, the identification of modified peptides by a bottom-up approach is not trivial. As an alternative approach and to overcome (some of) these problems, the interest for the analysis of proteins without any enzymatic digestion has tremendously grown the last few years [25]. This approach, referred to as top-down proteomics, allows the characterization of intact proteins from their MS/MS spectra [26,27]. Different proteoforms – all the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spliced RNA transcripts and PTMs [28] – can be identified localizing differences in their amino acid sequences or PTMs. Top-down analysis requires mass spectrometers with very high resolving powers and mass accuracy (*e.g.* Orbitrap®, Fourier transform ion cyclotron resonance (FTICR)). Unfortunately, ultrahigh resolution comes at the cost of longer analysis times, thus hampering application in traditional LC/MS/MS experiments. In general, ESI of an intact protein leads to highly charged ions with large isotopic distribution and different charge states. The fragmentation of intact proteins in an MS/MS experiment yields a large number of fragment ions, the isotopic distributions of which may overlap each other, thereby increasing the complexity of the spectra. In order to tackle this complexity, very high resolution measurements and consequent long experimental times are needed. Thus, the identification and

characterization of intact proteins is mainly performed by direct infusion- (DI) ESI-MS/MS analysis after off-line clean-up and/or fractionation procedures. However, the last generation Orbitrap® and high resolution time-of-flight (TOF) mass analysers offer the sensitivity, the mass resolution and the acquisition speed needed for top-down proteomics on a LC time-scale [29].

Tandem-MS or MS/MS of peptides and proteins can be performed in various ways using different fragmentation techniques [30]. Fragmentation by collision-induced dissociation (CID) occurs after acceleration of the ionized molecules by an electric field and collision with neutral gasses (*e.g.* helium, nitrogen or argon). Upon collision, part of the translational energy of the ionized molecules is converted into internal energy leading to the breakage of the peptide bonds and fragmentation of the precursor ion into smaller fragments. In a second and increasingly used dissociation process, electron-transfer dissociation (ETD) [31], low energy electrons are transferred from radical anions to multiply protonated molecules in the gas phase (*i.e.*  $[M+nH]^{n+}$ ) inducing random cleavages along the polypeptide backbone. Reactive radical anions are generated by reaction of neutral molecules of reagent (*e.g.* fluoranthene) and low energy electrons (*i.e.* lower than 1 eV) produced in a chemical ionization source after bombardment of a mediator gas (*e.g.* methane) with high energy electrons (*i.e.* 60-80 eV). The types of fragment ions obtained by CID and ETD depend on many factors including the amino acid sequence, the amount of internal energy, the charge state, the amount of radical anions, the reaction time, etc. As depicted in Figure 2, CID predominantly generates *b*- and *y*-type ions while ETD mainly leads to *c*- and *z*-type ions [32].



**Figure 2. Peptide fragment ions observed in MS/MS spectra. The type of fragments generated in a MS/MS experiments depend on used fragmentation technique. Collision-induced dissociation predominantly yields *b*- and *y*-type ions while electron transfer dissociation results in *c*- and *z*-type ions.**

One of major advantages of ETD in MS-based proteomics is the ability to localize the exact site of PTMs that can be missed by CID [33]. For example, CID of glycosylated peptides/proteins primarily leads to the fragmentation of the glycan structure with very low efficiency in the fragmentation of the polypeptide backbone. Instead, ETD primarily generates *c*- and *z*-type fragment ions without any loss of the glycosylation. This

property allows a better characterization of protein isoforms and PTMs, especially in top-down experiments [34]. It has been shown that CID and ETD provide complementary information in terms of the primary sequence coverage of unmodified peptides and that the combination of these two methods in bottom-up experiments improves the total number of identified species [35]. In electron-capture dissociation (ECD) the multiply protonated ions overlap with “free” high energy electrons (released from a cathode) that results in a type of fragmentation that is similar to ETD experiments [36,37]. This technique has been developed in FTICR-MS and is often applied for the characterization of intact proteins by top-down MS/MS.

The generation of multiply charged ions using ESI facilitates MS/MS-experiments, but on the other hand complicates the mass spectrum and thus limits the analysis of complex mixtures. On the contrary, MALDI of peptides and proteins generates mainly singly charged ions (*i.e.* mainly  $[M+H]^+$  or  $[M-H]^-$ ) thus yielding MS spectra with a reduced number of overlapping species that are relatively easy to interpret [38]. In MALDI, the analyte is co-crystallized with an excess of a small, organic chemical compound (*i.e.* chemical matrix) that adsorbs laser energy and distributes this energy to the sample. This process induces a very rapid heating of the sample generating the vaporization of both the matrix and the analyte with as a consequence the removal of a certain amount of the sample (the so-called ablation). In conjunction with the ablation process the ionization of a minor fraction of matrix and analyte occurs. With high-end MALDI sources proteins in the concentration range of as low as attomoles can be detected. Apart from other factors the ionization efficiency depends on the used chemical matrix. Alpha-cyano-4-hydroxycinnamic acid (CHCA) is a small aromatic acid particularly suited for the ionization of peptides and proteins up to 20 kDa.

CHCA crystallizes in a homogeneous manner and provides MALDI-spots that allow automated MS acquisitions. Surface enhanced laser desorption/ionization (SELDI) is a variation of MALDI that is based on a chemically modified surface to selectively capture peptides and proteins from a complex mixture [39]. The sample preparation procedure (*i.e.* capture and washing) is performed directly on a modified target plate improving the throughput of the analysis.

**Time-of-flight mass spectrometers.** In a TOF mass analyser, ions with different mass-to-charge ( $m/z$ ) ratios are separated in a flight tube based on their different flight times. To this end, *all* ions are accelerated to a defined kinetic energy (*i.e.*  $\frac{1}{2}mv^2$ ) and introduced into the flight tube *simultaneously* (delayed extraction). The accelerated ions then travel in a field-free region to a multichannel plate (MCP) detector. The “heavier ions” travel slower than the “lighter ions” and thus will be detected *sequentially*. By also measuring ions (molecules) with a known  $m/z$ -value the instrument can be calibrated through simple correlation of the flight time with the  $m/z$ -value. The mass resolution (*i.e.*  $\Delta\text{mass}/\text{mass}$ ) of a linear TOF mass analyser depends on the time interval in which ions with identical  $m/z$ -values reach the MCP detector. Factors that determine this interval are temporal distribution, spatial distribution and energy distribution of the ions. In modern mass analysers these factors are reduced through the implementation of specific technologies or configurations. For example, in reflectron TOF analysers the effect of the initial kinetic energy distribution on the mass resolution is counter-balanced using an electrostatic field for ion reflection. The reflectron, or ion mirror, thus allows compensating small ion energy differences by adjusting voltages. As a result, ions are more focused in time and mass resolution is improved, however at the cost of sensitivity for ions with higher  $m/z$ -

values, because in the reflectron configuration the transmission efficiency is decreased. Reflectron mode spectra are often referred to as high resolution profiles, with typical resolving powers of between 7,000 and 10,000 [40]. TOF instruments are “no scanning” and fast (*i.e.* low duty cycle) mass analysers that allow high-throughput measurements when combined with MALDI- or ESI sources. TOF analysers have been used in combination with both MALDI and ESI sources to analyse human serum peptides and proteins [41,42]. Linear MALDI-TOF MS allows the analysis of large intact proteins while reflectron MALDI-TOF MS is usually employed to analyse endogenous serum peptides or proteolytic fragments from larger proteins.

**Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometers.** The basis of an FTICR mass spectrometer consists of an analyzer (ICR) cell that resides in a strong, homogeneous magnetic field. At present, for MS-purposes magnets are commercially available with field strength up to 15 tesla, and development of 21T systems is on-going [43]. In an FTICR experiment, ions are trapped in an ICR cell using a front- and back-trapping electrode, followed by detection of their cyclotron frequency by two opposite detection electrodes in combination with two opposite excitation electrodes. Time schedules are crucial for proper control of any FTICR-MS or MS-MS experiment, consisting of ion quench/ion introduction, ion selection, ion dissociation, ion excitation, and ion detection [44,45].

Ion cyclotron movement, as a result of Lorentz force, is measured as a frequency that correlates with ion mass-to-charge ratio (*i.e.*,  $m/z$ ) and the field strength according to the following equation:

$$\omega_c = (qB_0) / m$$

In this initial condition, the ion cyclotron orbital phases are random and the cyclotron radius is too small to induce a detectable image current on the detection plates. Thus, the ions are accelerated to wider cyclotron orbits applying a uniform electric field (excitation), directed perpendicular to the magnetic field direction, and rotating at the cyclotron frequency (typically >10 kHz) of ions of a particular  $m/z$ -value (*i.e.* resonant). The final cyclotron radius is directly proportional to both the electric field strength and the time of the irradiation and is inversely proportional to the magnetic field strength. The excitation process leads to the spatial coherency of the ions making their detection possible. Ions of a wide  $m/z$ -range can be simultaneously excited using a broadband waveform. The image current generated after the excitation of the ions, decays in time and is generally referred to free induction decay (FID) transient. The FID resulting from the detection of ions with different  $m/z$ -values is a superposition of sine waves. At this point, Fourier transform is applied to resolve the FID and generate a frequency-domain spectrum that is further converted into a mass spectrum. The strength of the magnetic field affects many spectral characteristics. For example, the mass resolving power and mass measurement accuracy increase linearly with increasing field strength while the maximum number of ions that can be confined without coalescence of their FTICR signals increases quadratically.

FTICR analysers have been used in combination with both MALDI and ESI sources to analyse human serum peptides and proteins [46,47]. The main benefits of using this technology are (a) the ultrahigh mass resolving power that allows the analysis of large proteins and complex mixtures; (b) the ultrahigh mass measurement accuracy and



precision that allow a more reliable identification of the detected species; (c) a wide dynamic range that is favourable for the detection of low abundant components.

**Peptide and protein profiling of human serum samples.** Like other components in human serum, proteins may be used for the prognosis, diagnosis and monitoring of a certain disease (*e.g.* cancer, diabetes, atherosclerosis). In this context, both automated and high-throughput sample preparation methods and MS approaches have been developed to measure protein levels in biological samples with the analytical robustness required for clinical application [48]. Driven by the urgent need of biomarkers for the early detection of severe diseases such as cancer, MS is on its way to mature into a clinical diagnostic tool [49]. Numerous peptide and protein profiling studies by either SELDI- or MALDI-TOF-MS approaches have been reported for the discovery of disease-specific human serum signatures aiming for clinical diagnostics. Many of these studies were motivated by the early success published by Petricoin and co-workers in 2002. These investigators reported a serum proteomics pattern generated by RPC16 SELDI-TOF-MS that allowed the detection of ovarian cancer and the discrimination between healthy women and patients [50]. And indeed, over the last ten years, more disease-specific peptide and protein profiles related to other diseases could be identified through similar SELDI- or MALDI-TOF-MS methods [51,52]. However, also drawbacks of this approach became clear. First, only high-abundance serum proteins and their proteolytic fragments (*i.e.* peptides) were identified in the profiles [41]. Second, the identified discriminant proteins were not cancer-specific and in many cases appeared to be involved in the acute phase and inflammatory response [53,54]. Third, the standardization of sample collection and preparation procedures proved to be critical for the reproducibility of results.

As a consequence, none of proposed serum peptide signatures has been approved for routine diagnostics yet.

Due to the biological variability in human serum samples and the resulting proteomics profiles profiling studies require a high number of samples for analysis and evaluation to obtain reliable statistical results. Time-consuming or tedious sample preparation procedures (*e.g.* gel electrophoresis) and MS methods (*e.g.* LC-MS/MS) are not suitable for high-throughput analysis. Therefore, a robotic platform for sample clean-up and separation of complex protein mixtures is needed that provides a high degree of standardization and consequently low analytical variation. Solid-phase extraction (SPE) has been successfully implemented using ZipTip® pipette tips, cartridges or magnetic beads in automated sample preparation protocols for MS-based proteomics. In particular, the use of (disposable) magnetic beads (MBs)-based procedures leads to a fast and robust fractionation of human serum proteins. Proteins can be separated with high selectivity using different chemically functionalized beads such as reversed-phase material (RPC4, RPC8, RPC18). As an alternative, functionalized cartridges can be applied: these have a higher loadability (*i.e.* quantity of proteins that can be processed) thus providing larger fraction volumes suitable for further protein separation or (additional) direct infusion experiments.

Proteomic profiling studies are based on the comparison of hundreds (or even thousands) of signatures detected in hundreds of samples obtained from different individuals. After the acquisition of the MS profiles a series of data processing steps is required to extract the information needed for statistical analysis. These steps include a baseline correction, smoothing, peak-picking, internal calibration or peak alignment, and transformation of the data in a format compatible with methods for statistical analysis.

Data processing becomes more robust with improved quality of the profiles. Moreover, high quality data lead to a more reliable statistical analysis. In this context, resolving power, dynamic range, mass measurement accuracy and precision of the mass analyser are parameters that define the quality of the MS spectrum. The first proteomic profiling studies were performed on low resolution linear TOF mass analysers that detected peptides and proteins as broad peaks in a wide  $m/z$ -range. Later, high resolution reflectron mode TOF analysers were used to achieve isotopic resolution, however at a narrower  $m/z$ -range. Furthermore, the accuracy of SELDI/MALDI-TOF mass measurements is affected by the heterogeneity of the crystalized sample (*i.e.* chemical matrix and analytes) and both an external and an internal calibration (or alignment) are needed for robust comparisons of the profiles. Such differences in “height” of the sample on the MALDI target plate are less of a problem when applying MALDI-FTICR MS. Nevertheless, probably primarily due to the costs of this high-end platform, it has been rarely used for profiling of human serum peptides and proteins. However, following the latest developments of FTICR-MS (*e.g.* magnet production, more efficient ion transmission, cell improvements), the use of this technology has grown and the beneficial aspects have also become clear in clinical proteomics. Especially the ultrahigh resolution and the sub-ppm mass accuracy that are achieved with FTICR mass analysers have shown to be particularly useful for the identification of intact proteins and the localization of PTMs.

**Aim of this thesis.** In this thesis, the development and application of novel MALDI-FTICR-MS methods are described. These methods were used for the acquisition of ultrahigh resolution profiles containing (highly) abundant endogenous serum peptides and small proteins. It will be shown that the generation of high quality profiles by MALDI-

FTICR-MS results in a more robust and comprehensive analysis of the human serum degradome. This allows the identification of for instance cancer-related biomarker signatures, as is exemplified with serum samples from a pancreatic cancer cohort. The thus obtained proteomic signatures could be used for the accurate discrimination of healthy from diseased individuals. The developed MALDI-FTICR-MS methods were further applied to characterize small proteins fractionated by high-throughput and automated sample preparation protocols. Finally, identification of (large) polypeptides is reported using direct-infusion ESI-FTICR-MS/MS experiments.

## REFERENCES

1. Haddad, S. I.; Khairallah, A. A. A forgotten chapter in the history of the circulation of the blood. *Ann. Surg.* 1936, 104 (1), 1-8.
2. Hajdu, S. I. A note from history: The discovery of blood cells. *Ann. Clin Lab Sci.* 2003, 33 (2), 237-238.
3. Alfred Donné De l'origine des globules du sang de leur mode de formation et leur fin. *Comptes rendus de l'Académie des Sciences* 1842, (14), 366-368.
4. Gabriel Andral *Essai d'Hématologie Pathologique.* Fortin, Masson & Cie, Paris 1843.
5. William Addison *Experimental and Practical Researches on Inflammation and on the Origin and Nature of Tubercles of the Lung.* J Churchill, London 1843.
6. Zini, R.; Riant, P.; Barre, J.; Tillement, J. P. Disease-induced variations in plasma protein levels. Implications for drug dosage regimens (Part I). *Clin Pharmacokinet.* 1990, 19 (2), 147-159.
7. Zini, R.; Riant, P.; Barre, J.; Tillement, J. P. Disease-induced variations in plasma protein levels. Implications for drug dosage regimens (Part II). *Clin Pharmacokinet.* 1990, 19 (3), 218-229.
8. Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *PROTEOMICS* 2013, 13 (5), 723-726.
9. Tu, C.; Rudnick, P. A.; Martinez, M. Y.; Cheek, K. L.; Stein, S. E.; Slebos, R. J. C.; Liebler, D. C. Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics. *J. Proteome Res.* 2010, 9 (10), 4982-4991.
10. Hartley, P. The Determination of the Composition of the Different Proteins of Ox and Horse Serum by the Method of Van Slyke. *Biochem. J.* 1914, 8 (5), 541-552.
11. Rabilloud, T.; Carpentier, G.; Tarroux, P. Improvement and simplification of low-background silver staining of proteins by using sodium dithionite. *ELECTROPHORESIS* 1988, 9 (6), 288-291.
12. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* 1996, 68 (5), 850-858.
13. Håkansson, P.; Kamensky, I.; Sundqvist, B.; Fohlman, J.; Peterson, P.; McNeal, C. J.; Macfarlane, R. D. Iodine-127-plasma desorption mass spectrometry of insulin. *J. Am. Chem. Soc.* 1982, 104 (10), 2948-2949.
14. Heerma, W.; Kamerling, J. P.; Slotboom, A. J.; van Scharrenburg, G. J. M.; Green, B. N.; Lewis, I. A. S. Fast atom bombardment and collisional activation mass spectrometry of a heptapeptide. *Biol. Mass Spectrom.* 1983, 10 (1), 13-16.
15. Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 1984, 11 (11), 601.
16. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci.* 1986, 83 (17), 6233-6237.
17. Eng, J. K.; McCormack, A. L.; Yates III, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 1994, 5 (11), 976-989.

18. James, P. Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* 1997, 30 (4), 279-331.
19. Ferguson, P. L.; Smith, R. D. Proteome analysis by mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* 2003, 32, 399-424.
20. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 1993, 3 (6), 327-332.
21. Doerr, A. Mass spectrometry-based targeted proteomics. *Nat. Methods* 2013, 10 (1), 23.
22. Jmeian, Y.; El, R. Z. Liquid-phase-based separation systems for depletion, prefractionation and enrichment of proteins in biological fluids for in-depth proteomics analysis. *Electrophoresis* 2009, 30 (1), 249-261.
23. Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. Principles and practice of electrospray ionization-mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* 1991, 10 (5), 359-452.
24. Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Fröhlich, F.; Cox, J.; Mann, M. Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. *Mol. Cell. Proteomics* 2011, 10 (8).
25. Whitelegge, J. Intact protein mass spectrometry and top-down proteomics. *Expert. Rev. Proteomics* 2013, 10 (2), 127-129.
26. Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, 480 (7376), 254-258.
27. Ansong, C.; Wu, S.; Meng, D.; Liu, X.; Brewer, H. M.; Deatherage Kaiser, B. L.; Nakayasu, E. S.; Cort, J. R.; Pevzner, P.; Smith, R. D.; Heffron, F.; Adkins, J. N.; Pasa-Tolic, L. Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella Typhimurium* in response to infection-like conditions. *Proc. Natl. Acad. Sci. USA* 2013, 110 (25), 10153-10158.
28. Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* 2013, 10 (3), 186-187.
29. Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Müller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J. J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol. Cell. Proteomics* 2012, 11 (3).
30. Jones, A. W.; Cooper, H. J. Dissociation techniques in mass spectrometry-based proteomics. *Analyst* 2011, 136 (17), 3419-3429.
31. Kim, M. S.; Pandey, A. Electron transfer dissociation mass spectrometry in proteomics. *PROTEOMICS* 2012, 12 (4-5), 530-542.
32. Steen, H.; Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 2004, 5 (9), 699-711.
33. Mikesch, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* 2006, 1764 (12), 1811-1822.
34. Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer

Dissociation Mass Spectrometry. *Anal. Chem.* 2011, 83 (23), 8919-8927.

35. Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* 2011, 10 (5), 2377-2388.

36. Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization. *J. Am. Soc. Mass Spectrom.* 2012, 23 (10), 1707-1715.

37. Cooper, H. J.; Håkansson, K.; Marshall, A. G. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom. Rev.* 2005, 24 (2), 201-222.

38. Hillenkamp, F.; Karas, M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. In *Methods in Enzymology Mass Spectrometry*, Volume 193 ed.; James, A. M., Ed.; Academic Press: 1990; pp 280-295.

39. Merchant, M.; Weinberger, S. R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 2000, 21 (6), 1164-1177.

40. Nicolardi, S.; Palmblad, M.; Dalebout, H.; Bladergroen, M.; Tollenaar, R. A. E. M.; Deelder, A. M.; van der Burgt, Y. E. M. Quality control based on isotopic distributions for high-throughput

serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* 2011, 25 (23), 3457-3463.

MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spectrom.* 2010, 21 (9), 1515-1525.

41. Hortin, G. L. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.* 2006, 52 (7), 1223-1237.

42. Tucholska, M.; Florentinus, A.; Williams, D.; Marshall, J. G. The endogenous peptides of normal human serum extracted from the acetonitrile-insoluble precipitate using modified aqueous buffer with analysis by LC-ESI-Paul ion trap and Qq-TOF. *J. Proteomics* 2010, 73 (6), 1254-1269.

43. Kelleher, N.; Pa+ia-Toli-ç, L. 25th ASMS Sanibel Conference on Top Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2013, 24 (7), 983-985.

44. Amster, I. J. Fourier Transform Mass Spectrometry. *J. Mass Spectrom.* 1996, 31 (12), 1325-1337.

45. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* 1998, 17 (1), 1-35.

46. Palmblad, M.; van der Burgt, Y. E. M.; Mostovenko, E.; Dalebout, H.; Deelder, A. M. A novel mass spectrometry cluster for high-throughput quantitative proteomics. *J. Am. Soc. Mass Spectrom.* 2010, 21 (6), 1002-1011.

47. Nicolardi, S.; Palmblad, M.; Hensbergen, P. J.; Tollenaar, R. A. E. M.; Deelder, A. M.; van der Burgt, Y. E. M. Precision profiling and identification of human

48. Pham, T. V.; Piersma, S. R.; Oudgenoeg, G.; Jimenez, C. R. Label-free mass spectrometry-based proteomics for biomarker discovery and validation. *Expert. Rev. Mol. Diagn.* 2012, 12 (4), 343-359.

49. Palmblad, M.; Tiss, A.; Cramer, R. Mass spectrometry in clinical proteomics - from the present to the future. *PROTEOMICS - Clinical Applications* 2009, 3 (1), 6-17.
50. Petricoin III, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 2002, 359 (9306), 572-577.
51. Albrethsen, J. The first decade of MALDI protein profiling: a lesson in translational biomarker research. *J. Proteomics* 2011, 74 (6), 765-773.
52. Karpova, M. A.; Moshkovskii, S. A.; Toropygin, I. Y.; Archakov, A. I. Cancer-specific MALDI-TOF profiles of blood serum and plasma: Biological meaning and perspectives. *J. Proteomics* 2010, 73 (3), 537-551.
53. Le, L.; Chi, K.; Tyldesley, S.; Flibotte, S.; Diamond, D. L.; Kuzyk, M. A.; Sadar, M. D. Identification of Serum Amyloid A as a Biomarker to Distinguish Prostate Cancer Patients with Bone Lesions. *Clin. Chem.* 2005, 51 (4), 695-707.
54. Tolson, J.; Bogumil, R.; Brunst, E.; Beck, H.; Elsner, R.; Humeny, A.; Kratzin, H.; Deeg, M.; Kuczyk, M.; Mueller, G. A.; Mueller, C. A.; Flad, T. Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. *Lab. Invest.* 2004, 84 (7), 845-856.





# Chapter 1

Standardized and automated solid-phase  
extraction procedures for high-throughput  
proteomics of body fluids

*Bladergroen M. R., Derks R. J.E.,  
Nicolardi S., de Visser B., van Berloo S.,  
van der Burgt Y. E. M., Deelder A. M.*

Journal of Proteomics  
2012, 77, 144-153

## **ABSTRACT**

In order to balance the speed of analytical sample preparation procedures with mass spectrometry (MS)-based clinical proteomics the application of high-throughput robotic systems for body fluid workup is essential. In this paper we describe the implementation of various solid-phase extraction (SPE) sample preparation protocols on two different platforms, namely: 1) Magnetic bead-based SPE of peptides and proteins from body fluids on a Hamilton liquid handling workstation; 2) Cartridge-based SPE on a SPARK Symbiosis system. All SPE protocols were optimized for MS-based proteomics and compared with respect to obtained peptide- and protein profiles. Throughput numbers that were achieved in a 24 hour time frame for the sample workup procedures were more than 700 samples for the magnetic bead-based method and over 1000 samples for the cartridge-based method.

## INTRODUCTION

Proteomics has matured into a standard technology and with that interest has broadened from studying fundamental biological processes to screening large patient cohorts in clinical research [1]. Improved and more detailed genetic knowledge of various biological processes or organisms is now widely used in proteomics research [2;3]. The last decade sample throughput in MS-based proteomics has increased enormously due to the development of highly automated and robust mass spectrometers and improved speed in data handling and -processing. As a consequence, the analytical methods that involve peptide- and protein profiling of body fluids require high-throughput approaches. It should be stressed that sample workup is an essential part in all proteomics workflows because of the large complexity of any biological material [4-6]. In order to handle sample complexity and obtain full coverage multidimensional separations have been reported [7;8]. However, such approaches are not feasible for large screening studies taking into account time, costs and robustness. Most biomarker discovery studies that have reported on candidate peptides and proteins that were found to correlate with the presence and/or a stage of a certain disease were based on relatively small numbers of samples. Not surprisingly, a large majority of these discoveries could not be validated and did not make it into a diagnostic clinical assay [9;10]. Thus a high-capacity pipeline is needed to perform screening studies of clinical cohorts containing at least 1,000 different individuals [11]. Such a high-throughput workflow consists of robust and relatively fast sample processing and preparation (resulting in a reduced complexity) as well as automated MS [1;12]. Moreover, high-end MS in combination with specific sample workup allows mapping of modifications such as glycosylation and phosphorylation [13;14].

In this paper we present two automation platforms for SPE-based sample preparation and subsequent MS-measurements. Automation of sample preparation not only increases sample throughput but also improves robustness by eliminating human errors. Here, two different instrumental setups are described: The first one involves magnetic bead-based SPE, whereas the second SPE-setup is based on column-like cartridges. Both setups aim for peptide- and protein profiling of body fluids. For this purpose a subset of proteins and peptides is captured from body fluids such as serum or urine using an SPE approach. The captured components are measured on a mass spectrometer to obtain a “profile” (mass spectrum) of the body fluid [15]. Pattern analysis

of these profiles allows the classification of individuals into groups (*e.g.* healthy versus diseased). In general, these methods consist of incubation-, washing- and elution steps. In addition, a protocol is described that involves spotting of samples onto a MALDI-target plate for analysis on for instance a MALDI time-of-flight (TOF) or MALDI Fourier Transform Ion Cyclotron Resonance (FTICR) system. We have chosen to implement the first application on a Hamilton liquid-handling workstation. Advantages of this robotic system are the flexibility in deck layout, user-friendly programming software and the possibility to include in-house developed and third party consumables or equipment. The second application has been implemented on a SPARK Symbiosis SPE system, specifically adapted to increase sample throughput. The advantages of this system are the availability of a wide range of different cartridges and the ability to keep the samples at 4°C during sample storage and after elution.

The general availability of functionalized magnetic beads for peptide- or protein profiling is decreasing, or has even been discontinued in the case of (polymeric) reversed-phase (RP) C18-beads. Other players have entered the market by supplying magnetic beads based on silica material that are currently under investigation in our group. Furthermore, the magnetic bead protocol results in peptide- or protein eluates that are “optimized” for MALDI-MS measurements while in some instances direct infusion ESI-MS is the preferred MS method, *e.g.* to obtain multiply charged species that are more suitable for MS/MS-fragmentation. For these reasons an alternative SPE system was considered. This paper describes the implementation and comparison of both systems.

## **MATERIALS AND METHODS**

**Consumables and reagents.** All microtitration plates (PCR-plates) were obtained from Greiner (Alphen a/d Rijn, The Netherlands). Functionalized magnetic beads were either purchased from Bruker Daltonics (weak cation exchange (WCX); Bremen, Germany) or from Invitrogen (reversed-phase (RP) C18; Breda, The Netherlands). The elution buffer for WCX magnetic beads was a 130 mM ammoniumhydroxide solution (J.T. Baker, Deventer, The Netherlands) and the stabilization buffer consisted of 3% trifluoroacetic acid (TFA; Sigma, St. Louis, USA) in water. MALDI matrix  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) was obtained from Bruker Daltonics and used as a 3 mg/ml solution in

acetone:ethanol 1:2. Magnets were purchased from Webcraft GmbH (<http://www.supermagnete.de/>; Gottmadingen, Germany).

**Serum samples.** Serum samples were obtained from healthy volunteers at the Leiden University Medical Center (LUMC). Samples were collected, anonymized and processed according to a standardized protocol: all blood samples were drawn by antecubital venapuncture while the individuals were seated and had not been fasting. The samples were drawn in an 8.5 cc Serum Separator Vacutainer Tube (BD Diagnostics, Plymouth, UK) and within maximally 4 hours centrifuged at room temperature at 1000 g for 10 minutes. The samples were kept in sterile 500 µl barcode labeled polypropylene tubes (TrakMate, Matrix TechCorp.) at -80°C until further use.

**SPE using magnetic beads.** For each (biological) sample fresh magnetic beads were used. The RPC18-beads were first activated by a three-step washing with a 0.1% TFA solution. Then, the samples were added to the activated beads and incubated for 5 minutes at room temperature. The beads were washed again three times with 0.1% TFA and peptides were eluted with a 1:1 mixture of water and acetonitrile.

For the application of WCX-beads, the samples were added directly to the beads together with binding buffer since no activation was needed. The beads were washed three times with washing buffer and eluted with a high pH elution buffer, which was prepared as described in 'Consumables and reagents'. Thus obtained eluates were mixed with stabilization buffer to acidify the sample.

Two microliters of each (stabilized) eluate were mixed with 15 microliters of a MALDI matrix solution in a 384-well PCR plate. Then, 1 microliter of this mixture was spotted in quadruplicate onto a MALDI-target plate (Bruker Daltonics).

**SPE using cartridges.** For each (biological) sample a new cartridge was used. Cartridges were pre-washed with 1 ml 100% acetonitrile and equilibrated with 1 ml 1% acetic acid / 2% acetonitrile (wash solvent) prior to sample application. Serum samples were diluted 4 times with 0.1% acetic acid and 100 µL of the diluted serum was applied to the cartridge with 1 ml wash solvent. The cartridges were washed with 2 ml wash solvent and eluted

with 100  $\mu$ L 50% acetonitrile / 0.1% acetic acid. After elution the tubing was rinsed with 500  $\mu$ L wash solvent before proceeding to the next sample.

**Target plate spotting and mass spectrometry.** The protocol for spotting onto a MALDI-target plate is part of the other two procedures described above. The implementation will be described in the Results and Discussion section. MALDI-TOF experiments were performed on an UltraFlex II (Bruker Daltonics) either operating in positive reflectron mode in the  $m/z$ -range of 600–4,000 or in positive linear mode in the  $m/z$ -range of 1000–11000. The spectra were acquired using FlexControl software ver. 3.0 (Bruker Daltonics). A Smartbeam 200 Hz solidstate laser, set at a frequency of 100 Hz, was used for ionization. A profile, or summed spectrum, was obtained for each MALDI-spot by adding 20 spectra of 60 laser shots each at different rasters. To this end, FlexControl software decided on-the-fly whether or not a scan was used for the summed spectrum. For reflectron mode acquisitions, a resolution higher than 2,000 was required. Peaks were detected using the SNAP centroid peak detection algorithm with signal-to-noise threshold of 1 and a “TopHat” baseline subtraction. All mass scans not fitting these criteria were excluded. For linear mode acquisitions, a resolution higher than 100 was required and peaks were detected using the centroid peak detection algorithm with signal-to-noise threshold of 2 and a “TopHat” baseline subtraction. The measurement of a MALDI spot was finished when 1200 laser shots had been summed in one profile. FlexAnalysis Software 3.0 (Bruker Daltonics) was used for visualisation and data processing.

MALDI-FTICR experiments were performed on a Bruker 15 tesla solariX™ FTICR mass spectrometer equipped with a novel CombiSource [16]. The MALDI-FTICR system was controlled by Compass solariXcontrol software and equipped with a Bruker Smartbeam-II™ Laser System that operated at a frequency of 500 Hz. The “medium” predefined shot pattern was used for the irradiation. Each mass spectrum was obtained from a single scan of 600 laser shots using 512 K data points. Typically, the target plate offset was 100 V with the deflector plate set at 180 V. The ion funnels operated at 100 V and 6.0 V, respectively, with the skimmers at 15 V and 5 V. The trapping potentials were set at 0.60 V and 0.55 V, the analyzer entrance was maintained at -7 V, and side kick technology was used to further optimize peak shape and signal intensity. The required

excitation power was 28% with a pulse time of 20.0  $\mu$ s. MALDI-FTICR profiles were obtained from the same target plate that had been used for the MALDI-TOF acquisitions.

## RESULTS AND DISCUSSIONS

### **Implementation of SPE using magnetic beads on the robotic liquid handling system.**

The robotic system for automated liquid handling is built up from three units (Figure 1A). Currently, this system is used for (1) in-gel and in-solution digestions, (2) desalting of samples using ziptips, and (3) magnetic bead based SPE for peptide- and protein profiling. In this study data is reported from the third application, which involves the unit depicted on the right-hand side of Figure 1A. This unit consists of a two-arm Hamilton STARplus robotic workstation that is equipped with eight independently controlled 1 ml pipetting channels, a 300  $\mu$ L 96 channel pipetting head, an 8+1 channel nanospotter, a 96-well magnet, a vacuum system, a heated/cooled shaker, a temperature controlled carrier and a plate CORE-gripper. Description of the second and third unit (left-hand side and middle of Figure 1A) is outside the scope of this work.

For the application of magnetic bead based SPE it was necessary to develop specific labware that was not readily available from the supplier. For example, to prevent evaporation of volatile solvents from the supplied open reagent containers, special containers have been implemented with automated sliding lids to open and close on demand. In addition, a wireless and non-fixed mini-camera that can be picked-up by one of the eight 1 ml pipetting channels has been implemented for quality control purposes. Finally, a specific magnet plate incorporating 96 individual magnets to accommodate 96-well plates was in-house developed. With this magnet different sample workup steps can be performed as will be described in the following paragraph.

In MS-based clinical profiling studies proteins and peptides are isolated from a biological fluid (blood, urine, CSF) using SPE material, which is normally used in a cartridge or column-like format or as paramagnetic beads. In the here described automated system paramagnetic beads coated with a variety of functionalities, are applied to extract various subsets of peptides and proteins. The platform allows multiplexing, *i.e.* a sample plate can be processed with different types of magnetic beads, either simultaneously or sequentially. The deck layout for this protocol is shown in Figure 1B.



The various protocols for the different bead types have an identical workflow in terms of “order of events”, however slightly differ with regard to the number of activation cycles, wash cycles and incubation times and the specific buffers used. All protocols for the extraction procedures provided by the manufacturer for manual processing were adapted and further optimized to allow implementation on the robotic system. Initially, these optimizations have been described on a previous 8-channel liquid handling platform ([http://www.hamiltonrobotics.com/fileadmin/user\\_upload/proddb/app\\_notes/Proteomics/BR-0504-01\\_PeptideExtraction\\_for\\_ProteinProfiling\\_01.pdf](http://www.hamiltonrobotics.com/fileadmin/user_upload/proddb/app_notes/Proteomics/BR-0504-01_PeptideExtraction_for_ProteinProfiling_01.pdf)). However, for the current platform these protocols were further adjusted to obtain an even higher throughput. In short, each protocol consists of a binding step, several washing steps and an elution step. Optionally there might be an activation and/or equilibration step, and a stabilization step of the eluate. All eluates are collected in a fresh PCR-plate and then spotted in quadruplicate onto a target plate for further analysis by MALDI-TOF-MS and/or MALDI-FTICR-MS. In order to reduce loss of magnetic beads to a minimum PCR-plates were used throughout the protocol. These plates allow for maximum aspiration of solvents with a minimal residual volume in the wells and minimal disturbance of the magnetic bead pellet through a smaller surface to height ratio in comparison with other types of microtitration plates.

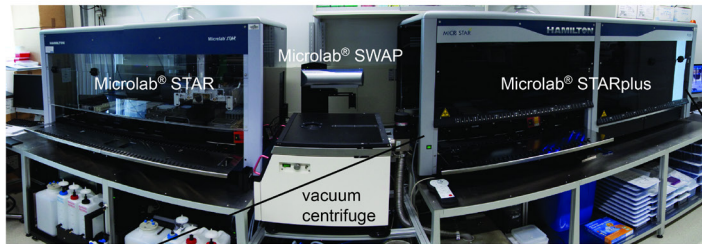
Obviously, the manipulation of magnetic beads requires a magnet for capturing. Ideally, in a 96-well PCR plate all magnetic beads are captured at the bottom of each well upon using a magnet. Different types of commercially available magnets to accommodate such 96-well plates were tested on the Hamilton system, namely magnets in ring-, bar- and pin form. Unfortunately, in our hands none of these magnets allowed suitable implementation in the robotic system due to “off-center” or “well-above-the-bottom” capturing of magnetic beads in a PCR plate. As a result of non-optimal capture proper resuspension of magnetic beads was not possible and the SPE-cleanup of serum samples failed. Therefore, a magnet block was in-house developed with a single magnet at the bottom of each well, resulting in optimal capture of beads. Proper resuspension of the bead pellet needs to be performed without the magnet. This implicates that two deck positions are required for processing one sample plate. Therefore, an electronically controlled height-adjustable magnet was designed to allow collection and resuspension of beads at the same position in the robot deck.

The here described SPE-protocols based on magnetic bead fractionation were implemented in the year 2008. Since then, more than 15,000 different serum samples have been processed in biomarker screening studies [17-23]. These studies included statistical validation tests. From the results of the 96-channel system it was concluded that the reproducibility of the peptide- and protein profiles obtained from MALDI-TOF-MS was better than in the case of manual sample processing.

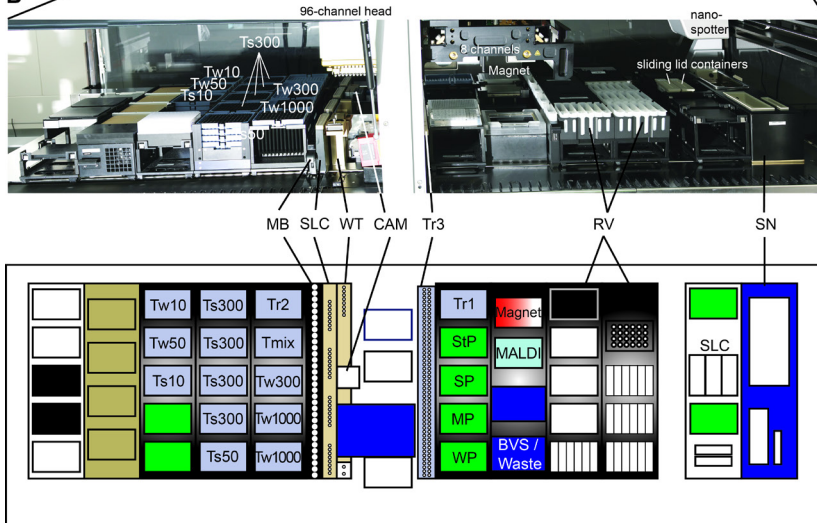
Typically, the coefficient of variation (CV) of the peak areas in the profiles was on average 15% when processing biological serum sample replicates, and for some specific peptides even below 5%. Note that this value results from a combination of variation in the SPE-procedure itself, the MALDI-spotting and MALDI-TOF measurement. This aspect is further discussed in the last paragraph of this section (comparison of beads and cartridges).

**Implementation of online SPE using cartridges on the Symbiosis.** Basically, a Symbiosis system consists of one autosampler (Reliance), one High Pressure Dispenser (HPD) and one Automated Cartridge Exchange (ACE) unit. In Figure 1C, the setup of our customized Symbiosis system is shown. In this setup three Reliance systems are combined with two HPD units, each equipped with two syringes, and two ACE units. Two Reliance systems are used as an autosampler (left- and right-hand side of Figure 1C) and the third Reliance acts as a fraction collector (middle of Figure 1C). The fraction collector is reached by both systems in an alternating way: when one system is in the “load position”, *i.e.* equilibrating and washing a cartridge (left-hand side in Figure 1C), the other one is in the “elute position” (right-hand side in Figure 1C). Note that the cartridge is swapped from the left clamp to the right clamp in the ACE unit on the left-hand side upon elution, and vice-versa for the cartridge in the second ACE. Moreover, for each sample a new cartridge is applied, and the used cartridge is placed back into the storage container. In this approach a throughput of about 45 samples per hour can be achieved.

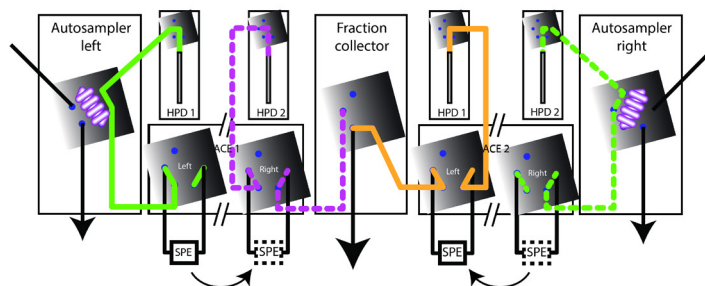
A



B



C

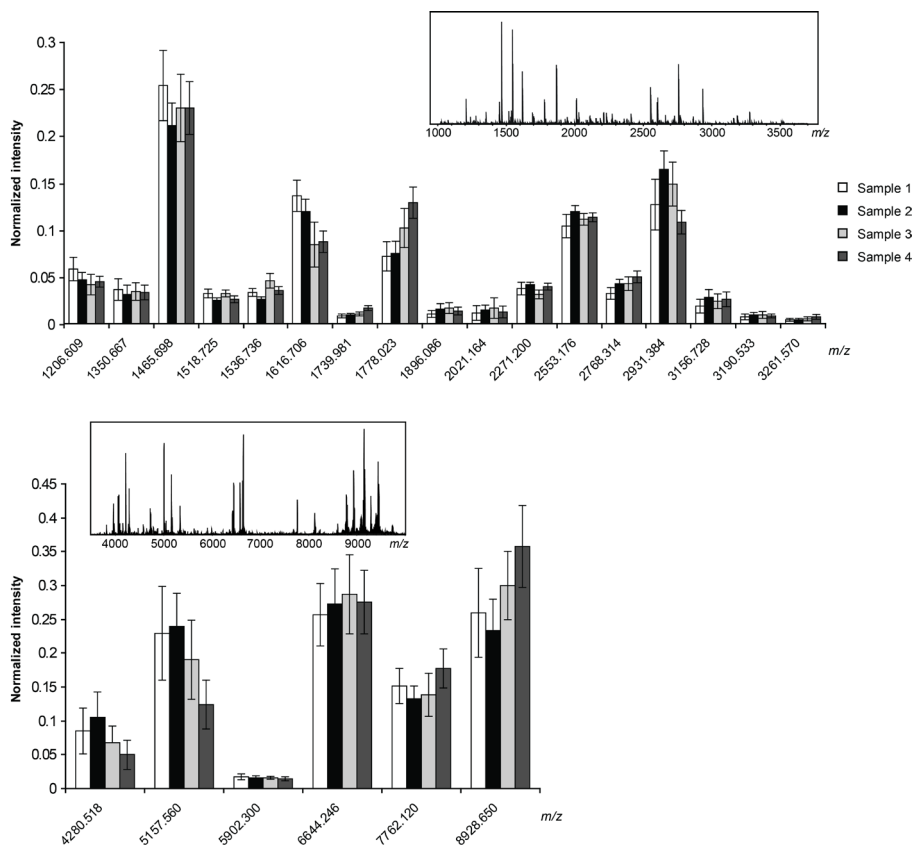


**Figure 1. Systems setup. Picture of both sample preparation systems, consisting of a Hamilton Microlab<sup>®</sup> STAR, a Hamilton SWAP, an automated CHRIST vacuum centrifuge and a Hamilton Microlab<sup>®</sup> STARplus (A), a photo and schematic representation (B) of the Microlab<sup>®</sup> STARplus deck layout used for magnetic bead based protein profiling and a photo with flow-scheme (C) of the Symbiosis system. To carry out a protocol on the Microlab<sup>®</sup> STARplus with optimal use of deck space and without user-intervention stacks of pipette tips are needed. This storage of tips on the deck is further indicated as “Tip-stacking” (Ts) positions. To allow for tip pickup within a certain protocol one rack of tips is initially transferred to a so-called “Tip-working” (Tw) position. TCC: Temperature Controlled Carrier; Tw10: tip work sequence 10 microliter tips; Tw50: tip work sequence 50 microliter tips; Tw300: tip work sequence 300 microliter tips; Tw1000: tip work sequence 1000 microliter tips; Ts10: tip stack 10 microliter tips; Ts50: tip stack 50 microliter tips; Ts300: tip stack 300 microliter Tips; Tmix: spare position for all tip types to be used by 96-channel head; Tr1, Tr2, Tr3: empty racks for tips to be re-used; MB: Magnetic beads; WP: working plate; MP: mixing plate for mixing sample and MALDI-matrix; SP: sample plate; StP: eluent storage plate; SLC: Sliding Lid Containers for volatile solvents; MALDI: MALDI-target plate; RV: reagent vessels; WT: waste and tool carrier. Tools include CORE-gripper and CORE-camera (CAM); SN: service carrier nano-head for washing and waste.**

Both paramagnetic beads and the cartridges are available with a wide variety of functionalities. However, cartridges can also be specifically packed with any solid phase material of choice, making the availability of such material the only limiting factor. In this study six out of eight functionalities from the SPARK method development kit were tested with respect to peptide- and protein isolation from serum samples, as well as three other functionalities, namely: CN-SE, C2-SE, C8 EC-SE (End Capped), C18 HD (High Density), Resin GP (General Phase), Resin SH (Strong Hydrophobic), PLRP, C18-OH and C18-EC. A general protocol was used for all cartridges. In short, the cartridges were activated with acetonitrile and equilibrated with wash solvent. Serum samples were diluted 4 times with 0.1% acetic acid and applied to the cartridges using wash solvent. The cartridges were washed and eluted with elution solvent. The eluates were collected as 100  $\mu$ L fractions. The resulting eluates were spotted in quadruplicate using the Hamilton STARplus as described below.

Figure 2 (inset) shows a typical FTICR-MS profile of SPE eluates, such as obtained with samples processed with the nine cartridges mentioned above. Based on the number of features (peaks) in the spectrum and the total signal intensity in both the low- and the high-mass range, the C8 EC-SE, C18-HD and C18-EC cartridges performed best and to a lesser extent also the C2-SE in the high-mass range. Preliminary results also indicated that the eluates from the C18-EC cartridges performed best in direct infusion FTICR-MS experiments, making these cartridges the type of choice for future experiments.

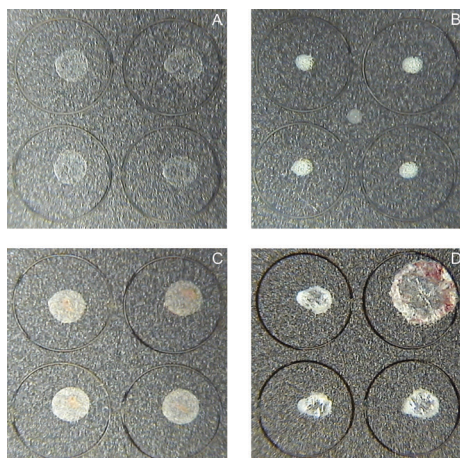
Using these C18-EC cartridges, four samples were processed in eightfold and spotted in quadruplicate onto a MALDI-target plate. For each processed sample two spots were used for MALDI-FTICR-MS measurements in the low mass range (1000-3700 Da) and two spots were used for MALDI-FTICR-MS measurements in the high mass range (3.5-10kDa), resulting in 16 spectra for each sample in both mass ranges. For 23 peaks of known composition and present in all spectra [16;24], the intensities expressed as area under the curve were averaged over the 16 spectra. The intensities were normalized against the total intensity of the 17 (low-mass range) and 6 (high-mass range) peaks, respectively. The results are depicted in Figure 2. It can be seen that the intra-sample reproducibility (standard deviation) is dependent on the peptide itself (*i.e.* the  $m/z$ -value), and in some cases this error is even larger than the inter-sample variation (*e.g.* at  $m/z$ -value 1465). However, in general the standard deviations are within acceptable limits and allow accurate comparison between peptide intensities in different serum samples. For example, the peptide intensity at  $m/z = 1778.023$  is two-fold higher in sample 4 compared to sample 1.



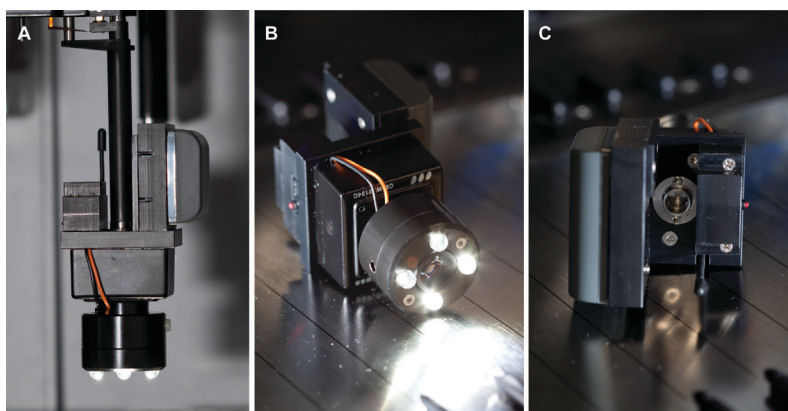
**Figure 2. Repeatability and Reproducibility of the cartridge-based SPE method. Peptides used for this analysis are observed as  $[M+H]^+$  at monoisotopic  $m/z$  values:** 1206.6 = EGDFLAEGGGVR; 1350.6 = SGEGDFLAEGGGVR; 1465.7 = DSGEGDFLAEGGGVR; 1518.7 = ADS(-H<sub>2</sub>O)GEGDFLAEGGGVR; 1536.7 = ADSGEGDFLAEGGGVR; 1616.7 = ADSpGEGDFLAEGGG VR; 1739.9 = NGFKSHALQLNRRQI; 1778.0 = SKITHRIHWESASLL; 1896.0 = RNGFKSHALQLN NRQI; 2021.1 = SSKITHRIHWESASLLR; 2271.1 = SRQLGLPAPPDVPDHAAYHPF; 2553.1 = SSSYSKQFTSSTSYPNRGDSTFES; 2768.2 = SSSYSKQFTSSTSYPNRGDSTFESKS; 2931.3 = SSSYSKQFTSSTSYPNR GDSTFESKSY; 3156.6 = NVHSGSTFFKYLLQAKIPKPEASFSPR; 3190.4 = SSSYSKQFTSSTSYPNRGDSTFESKSYKM; 3261.5 = SSSYSKQFTSSTSYPNRGDSTFESKSY KMA; 4280.5= NVHSAGAAGSRMNFPRGVLSSRQLGLPAPPDVPDHAAYHPF; 5902.4 = SSSYSKQFTSSTSYPNRGDSTFESKSYKMADEAGSEADHEGTHSTKRGHAKSRPV; 6644.3 = PSPTFLTQVKESLSSYWEAKTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMoxSTYTGIFT; 8928.8 = TQQPQQDEMOxPSPTFLTQVKESLSSYWEAKTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMSTYTGIFTDQVLSVLKGEE

**Implementation of a MALDI-target plate spotting protocol on the robotic liquid handling system.** A spotting protocol for MALDI plates has been developed previously ([http://www.hamiltonrobotics.com/fileadmin/user\\_upload/prodb/app\\_notes/Proteomics/BR-0302-01\\_MALDITargetSpot.pdf](http://www.hamiltonrobotics.com/fileadmin/user_upload/prodb/app_notes/Proteomics/BR-0302-01_MALDITargetSpot.pdf)); however this protocol included a washing step on the MALDI target. Since our procedures result in purified samples such a washing step was not needed anymore. To this end, we compared two methods of matrix and sample deposition on the MALDI-target: 1) Premixing sample with matrix in a 384-well plate and 2) deposition of sample on a MALDI target plate followed by application of matrix with the nanospotter. The two methods differ in sample to matrix ratio and the final concentration of organic solvent. For the first method the sample to matrix ratio was 2:15 and the organic solvent concentration was 94% while for the second method the sample to matrix ratio was 10:4 and the organic solvent concentration was 64%, when calculated for a sample composition containing 50% acetonitrile as is the case for most of the applications described in this paper. This ratio is of influence on the drying speed and crystal formation. The premixing method was found to perform consistently better than the nanospotter method. In principle it would be possible to reverse the second method (apply sample instead of matrix with the nanospotter). This would change the sample to matrix ratio and the final organic solvent concentration, but processing times would increase dramatically. Moreover, to prevent sample carry-over disposable tips are to be preferred over washable tips. Depending on the method applied before spotting, using only a single spot for each sample will not always result in a good spectrum.

Using  $\alpha$ -cyanohydroxycinnamic acid (CHCA) as the MALDI matrix and a Bruker Anchorchip<sup>TM</sup> an optimal spot is a very thin uniform layer which is slightly larger than the anchor (Figure 3A). Causes for non-ideal spots could *e.g* be the absence of analytes in the sample or extract, resulting in very dense spots with the size of the anchor (Figure 3B) or, in case of magnetic bead purification, the presence of magnetic beads on the MALDI target caused by incomplete separation of beads from the eluate, resulting in brown spots (Figure 3C). Furthermore, aberrant spot volumes, caused for instance by droplets at the outside of the tips, may result in different crystal formation and subsequently worse spectra (Figure 3D). Therefore, each sample is generally spotted multiple times.



**Figure 3.** Examples of MALDI spots. (A) Optimal spot: thin homogeneous matrix layer, which is larger than anchor. (B) Small spots almost the size of the anchor, due to absence or (very) low concentration of analytes. (C) Brown spots: presence of magnetic beads due to incomplete separation of magnetic beads from eluate. (D) One very large spot due to the presence of a droplet on the tip during spotting. Also visible are magnetic beads (brown color) and an inhomogeneous matrix layer due to changed evaporation conditions (lower organic solvent concentration).

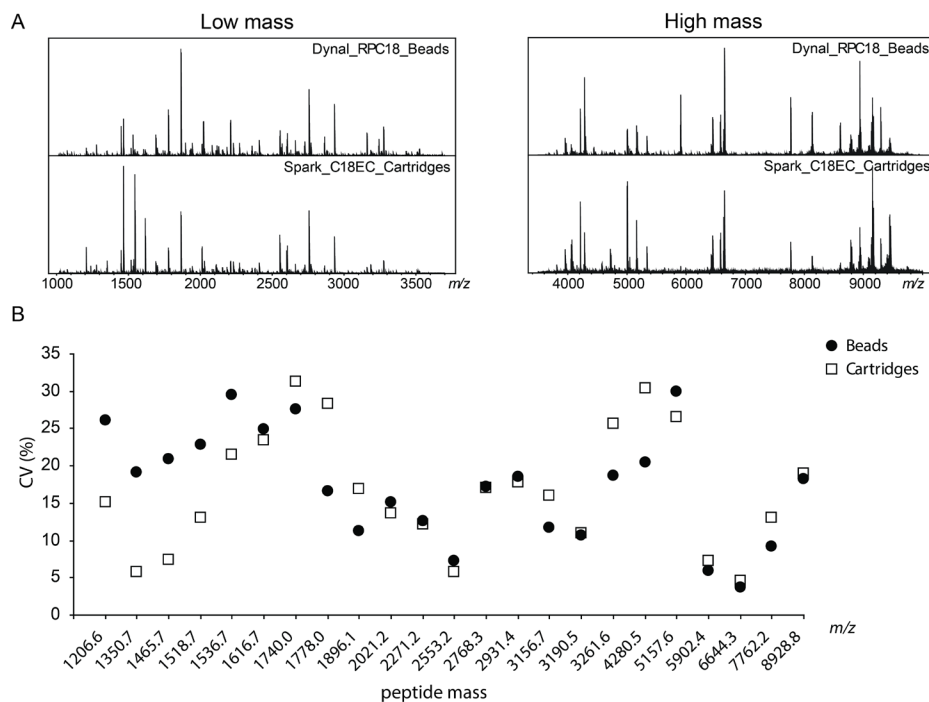


**Figure 4.** Wireless camera mounted on a pipetting channel (A) top-view showing the CORE-head, battery and switch (B) and bottom view showing the LEDs (C).



Theoretically it would be possible to check the quality of the spots and the envisaged quality of the spectra by imaging and automated evaluation of the images. For this purpose it would be possible to use a camera mounted on the Hamilton STARplus, as described in an application note on the Hamilton website ([http://www.hamiltonrobotics.com/fileadmin/user\\_upload/proddb/app\\_notes/Genomics/BR-0506-01-EasyPick\\_ColonyPicking.pdf](http://www.hamiltonrobotics.com/fileadmin/user_upload/proddb/app_notes/Genomics/BR-0506-01-EasyPick_ColonyPicking.pdf)). However, this is a serious investment. Therefore, an affordable system was developed which, in addition, is not fixed to the robot, but is placed on the deck as separate labware. To implement this, a commercially available wireless mini camera was mounted in a frame, together with a battery and LEDs for illumination. To grab the camera the Hamilton's CORE technology was used. For this purpose a CORE head, similar to a tip-head, was also included in the frame. Furthermore, a switch was integrated which activates the camera when this is picked up and stops it when placed back on the deck (Figure 4). So far, no software is available which allows automated image evaluation, but we foresee that general software for this purpose will become available in the future.

**Comparison of MALDI-FTICR profiles obtained from magnetic beads-based SPE or cartridge-based SPE.** The same samples from the reproducibility test described above were also processed with C18 magnetic beads. Here, the comparison of the bead-processed samples with the cartridge-processed samples is based on MALDI-FTICR profiles, although similar profiles were obtained using MALDI-TOF-MS (data not shown). Recently, we have reported on the beneficial effects of ultrahigh resolution MS in terms of determined mass measurement errors and alignment of data [16]. Typical examples of these precision profiles are depicted in Figure 5a, in which a so-called low mass and high mass spectra are shown. The comparison of eluates from C18 magnetic beads to those of C18-EC cartridges is illustrated in Figure 5b, where peak area CVs (four samples, eight independent workups each) are given for a set of known peptides (same as in Figure 2). At a first glance, especially the low mass region seems different between the two SPE protocols. However, upon careful analysis it was found that on average the same peptides are captured by both the cartridges and the magnetic beads, although for some peptides in a different ratio. For instance,  $m/z$  1465.7 and 1616.7 were captured to a higher



**Figure 5. Comparison of magnetic bead-based SPE with cartridge-based SPE. Spectra (A) and average normalized intensities (B). Peptides used for this analysis are the same as in Figure 2.**

extent by the cartridges, while  $m/z$  2021.2 and 3156.7 were isolated in higher amounts by the magnetic beads. Similarly, in the high mass region differences were observed for the signal at  $m/z$  5905.4. From Figure 5b it becomes clear that the peak area CVs are often similar for both the magnetic beads and the cartridges. Note that for some of the peptides the two workup protocols yield results with different efficiency and reproducibility. For these peptides the less efficient method usually exhibits the largest variation. It is obvious that biologically relevant differences should be considered with regard to the determined CV of each specific peptide.

In summary, precision profiles obtained from the same sample after work-up using the here described two SPE-protocols show a large amount of overlap in a qualitative analysis, *i.e.* the same peptides are captured by both procedures. From these results we conclude that the C18 cartridges provide a suitable alternative for C18

magnetic beads. Nevertheless, for some species large differences in the absolute amounts (ratios) are observed. Whether or not these intensity differences have an effect on classification of samples of specific patient cohorts is currently under investigation.

## **CONCLUSIONS**

The implementation of the sample preparation processes on automated liquid handling platforms, as described in this paper, allows to both speed up the analysis of a large number of samples but also to eliminate errors. Apart from the throughput, the latter aspect is essential for application in clinical diagnosis where standardization and quality control (QC) are pivotal in meeting good laboratory practice (GLP) requirements [25]. In combination with the previously described automated MALDI-target loading and the high-throughput MS-based methods developed in our group this workflow is a next step towards implementation of proteomics in a clinical setting [26;27].

In this paper, two sample preparation techniques based on SPE on a highly automated platform are described. It was found that the cartridge-based Symbiosis system is a suitable alternative for the magnetic bead-based sample preparation and cleanup. In view of the fact that the production of specific types of beads has been discontinued this is an important aspect. Furthermore, the use of cartridges increases flexibility in sample preparation, since the cartridges can be packed with virtually any desired solid phase material. The peptides captured with C18-SPE material by both techniques are similar, although the capture efficiency may differ for each peptide. Furthermore, it was shown that the standard deviations for both the magnetic beads and the cartridges were very similar in inter- as well as intra-sample comparisons. As discussed in this paper, the CV of the intensity of a certain peak in magnetic bead profiles was often well below 10% when processing biological serum sample replicates, and on average about 15%. Based on the results presented here we expect this to be the same for the cartridges, which requires further investigation with larger sample cohorts.

## **SUPPLEMENTAL INFORMATION**

A movie of the CORE-Camera in action can be found on <http://www.lumc.nl/con/1040>, then navigate to ‘biomolecular mass spectrometry/Mass spectrometry-based clinical proteomics/methods’.

## **ACKNOWLEDGEMENTS**

The authors wish to thank Dr. O. Klychnikov from the Leiden University Medical Center department of Parasitology, Biomolecular Mass Spectrometry Unit, for assistance with digital imaging.

## REFERENCES

- [1] Nilsson T, Mann M, Aebersold R, Yates JR, III, Bairoch A, Bergeron JJ. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods* 2010;7:681-5.
- [2] Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de LU, Rinner O, Lee H, Pedrioli PG, Malmstrom J, Koehler K, Schrimpf S, Krijgsveld J, Kregenow F, Heck AJ, Hafen E, Schlapbach R, Aebersold R. A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat Biotechnol* 2007;25:576-83.
- [3] Nedelkov D, Kiernan UA, Niederkofler EE, Tubbs KA, Nelson RW. Investigating diversity in human plasma proteins. *Proc Natl Acad Sci U S A* 2005;102:10852-7.
- [4] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845-67.
- [5] Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 2006;312:212-7.
- [6] Beretta L. Proteomics from the clinical perspective: many hopes and much debate. *Nat Methods* 2007;4:785-6.
- [7] Jacobs JM, Adkins JN, Qian WJ, Liu T, Shen Y, Camp DG, Smith RD. Utilizing human blood plasma for proteomic biomarker discovery. *Journal of Proteome Research* 2005;4:1073-85.
- [8] 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.
- [9] Mitchell P. Proteomics retrenches. *Nat Biotechnol* 2010;28:665-70.
- [10] Surinova S, Schiess R, Huttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the development of plasma protein biomarkers. *J Proteome Res* 2011;10:5-16.
- [11] Anderson NL. Counting the proteins in plasma. *Clin Chem* 2010;56:1775-6.
- [12] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198-207.
- [13] Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 2004;76:3935-43.
- [14] Wuhrer M, Koeleman CA, Hokke CH, Deelder AM. Protein glycosylation analyzed by normal-phase nano-liquid chromatography--mass spectrometry of glycopeptides. *Anal Chem* 2005;77:886-94.
- [15] Albrethsen J. The first decade of MALDI protein profiling: a lesson in translational biomarker research. *J Proteomics* 2011;74:765-73.
- [16] Nicolardi S, Palmblad M, Hensbergen PJ, Tollenaar RA, Deelder AM, van der Burgt YE. Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 2011;25:3457-63.
- [17] Alagaratnam S, Mertens BJ, Dalebout JC, Deelder AM, van Ommen GJ, den Dunnen JT, 't Hoen PA. Serum protein profiling in mice: identification of Factor XIIIa as a potential biomarker for muscular dystrophy. *Proteomics* 2008;8:1552-63.
- [18] Nicolardi S, Palmblad M, Dalebout H, Bladergroen M, Tollenaar RA, Deelder AM, van der Burgt YE. Quality Control Based on Isotopic Distributions for High-Throughput MALDI-TOF

and MALDI-FTICR Serum Peptide Profiling. *J Am Soc Mass Spectrom* 2010.

[19] Huijbers A, Velstra B, Dekker TJ, Mesker WE, van der Burgt YE, Mertens BJ, Deelder AM, Tollenaar RA. Proteomic serum biomarkers and their potential application in cancer screening programs. *Int J Mol Sci* 2010;11:4175-93.

[20] Huijbers A, Mesker W, Velstra B, van Der Burgt Y, Mertens B, Deelder A, Tollenaar R. Early Detection of Colorectal Cancer Using Mass Spectrometry Based Serum Protein Profiling. *Ann Oncol* 2010;21:209.

[21] Velstra B, Mesker WE, van der Burgt YEM, Mertens BJ, Deelder AM, Tollenaar RAEM. Mass spectrometry based serum protein profiling for the early detection of breast cancer: Taking the steps towards clinical implementation. *Brit J Surg* 2010;97:S5.

[22] Mertens BJA, van der Burgt YEM, Velstra B, Mesker WE, Tollenaar RAEM, Deelder AM. On the use of double cross-validation for the combination of proteomic mass spectral data for enhanced diagnosis and prediction. *Stat Probabil Lett* 2011;81:759-66.

[23] Nadarajah VD, Mertens BJA, Dalebout H, Bladergroen MR, Alagaratnam S, Garrood P,

Bushby K, Straub V, Deelder AM, den Dunnen JT, van Ommen G-JB, 't Hoen PAC, van der Burgt YEM. Serum Peptide Profiles Of Duchenne Muscular Dystrophy (DMD) Patients Evaluated Using Data Handling Strategies For High Resolution Content. *J Proteomics Bioinform* 2012;4:96-103.

[24] Tiss A, Smith C, Menon U, Jacobs I, Timms JF, Cramer R. A well-characterised peak identification list of MALDI MS profile peaks for human blood serum. *Proteomics* 2010;10:3388-92.

[25] Lim MD, Dickherber A, Compton CC. Before you analyze a human specimen, think quality, variability, and bias. *Anal Chem* 2011;83:8-13.

[26] McDonnell LA, van Remoortere A., van Zeijl RJ, Dalebout H, Bladergroen MR, Deelder AM. Automated imaging MS: Toward high throughput imaging mass spectrometry. *J Proteomics* 2010;73:1279-82.

[27] Palmblad M, van der Burgt YE, Mostovenko E, Dalebout H, Deelder AM. A novel mass spectrometry cluster for high-throughput quantitative proteomics. *J Am Soc Mass Spectrom* 2010;21:1002-11.



# Chapter 2

Precision profiling and identification of  
human serum peptides using Fourier transform  
ion cyclotron resonance mass spectrometry

*Simone Nicolardi, Magnus Palmblad,  
Paul J. Hensbergen, Rob A. E. M. Tollenaar,  
André M. Deelder and Yuri E. M. van der Burgt*

Rapid Commun. Mass Spectrom.  
2011, 25, 3457–3463



## ABSTRACT

Many biomarker discovery studies are based on matrix-assisted laser desorption/ionisation (MALDI) peptide profiles. In this study, 96 human serum samples were analysed on a Bruker solariX™ MALDI Fourier transform ion cyclotron resonance (FTICR) system equipped with a 15 tesla magnet. Isotopically resolved peptides were observed in ultrahigh resolution FTICR profiles up to  $m/z$  6500 with mass measurement errors (MMEs) of previously identified peptides at a sub-ppm level. For comparison with our previous platform for peptide profile mass analysis (*i.e.* Ultraflex II) the corresponding time-of-flight (TOF) spectra were obtained with isotopically resolved peptides up to  $m/z$  3500. The FTICR and TOF systems performed rather similar with respect to the repeatability of the signal intensities. However, the mass measurement precision improved at least 10-fold in ultrahigh resolution data and thus simplified spectral alignment necessary for robust and quantitatively precise comparisons of profiles in large-scale clinical studies. From each single MALDI-FTICR spectrum an  $m/z$ -list was obtained with sub-ppm precision for all different species, which is beneficial for identification purposes and interlaboratory comparisons. Furthermore, the FTICR system allowed new peptide identifications from collision-induced dissociation (CID) spectra using direct infusion of reversed-phase (RP) C18-fractionated serum samples on an electrospray ionisation (ESI) source.

## INTRODUCTION

Mass spectrometry (MS) has emerged as a key technology for molecular profiling of biological samples. Nowadays this versatile and high-end analytical tool is applied in (pre)clinical studies to identify and quantify proteins, lipids, or metabolites [1–3]. The first strategy is commonly referred to as MS-based proteomics and is widely applied to identify and quantify proteins in for instance studies on cancer biology [4–6]. One way of performing an MS-based proteomics experiment is to map a set of (poly)peptides in a single spectrum, *i.e.* a profile. The continuous efforts to find new prognostic or diagnostic biomarkers have stimulated the use of such profiles in clinical settings. In the early days, a single low-resolution mass spectrum derived from an individual's body fluid was used for comparative studies. However, a profile of a complex mixture is most informative when recorded on an ultrahigh resolution instrument such as a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer, as was early recognised for the analysis of peptides, glycans or even crude oil [7–11]. Identifications of compounds are also more confident from ultrahigh resolution profiles. Unfortunately, it soon became clear that profiles reflected only a subset of the complex mixture and that online chromatographic separation in combination with electrospray ionisation (ESI)-MS would allow the samples to be analysed in greater depth [12,13]. Profiling studies became unpopular and were criticised because of their problems with reproducibility and validation [14]. In this context, the importance of standardisation in sample workup, measurement and data processing and evaluation has been stressed [15–17]. Thus, MS profiles can be highly reproducible provided each procedure in the workflow is carefully and correctly applied [1]. In addition, matrix-assisted laser desorption/ionisation (MALDI)-MS of peptides remains unrivalled with respect to acquisition speed and automation strategies [18]. Following the current developments in ultrahigh resolution MS, mass measurement at higher precision improves the alignment of multiple spectra that are needed for comparative analysis [19,20]. Here we demonstrate a robust method for peptide profiling using a commercially available MALDI-FTICR system equipped with a novel front-end and a 15 tesla magnet. This work builds on recent developments of MS systems and measurement strategies to meet the challenge of complexity inherent to biological samples. The high-mass measurement precision significantly aids data analysis workflows in discovery studies and is a key step towards robust and standardised high-throughput

screening of hundreds to thousands of patient samples with high reproducibility in a realistic timeframe.

## **EXPERIMENTAL**

**Sample collection and serum peptide isolation.** The protocols for both the collection of human blood samples and for serum peptide isolation have been described previously [17]. Informed consent was obtained from all patients and the Leiden University Medical Center (LUMC) Medical Ethical Committee approved the studies. Briefly, blood samples were collected in 10 mL BD Vacutainer tubes (containing a clot activator and a gel for serum separation) from patients and healthy volunteers by antecubital venipuncture and centrifuged within 1 h. Serum was then transferred into a 1 mL cryovial and stored at -80 °C until further aliquotting. To this end, 96 cryovials were thawed and serum was distributed over eight racks using an eight-channel liquid-handling robot (Hamilton, Bonaduz, Switzerland). Each rack was stored again at -80 °C and thawed only for the automated peptide isolation procedure. Serum peptides were isolated in a fully automated fashion using reversed-phase (RP)C18-functionalised magnetic beads. In this study, 10 mL of commercially available RPC18 Dynabeads (Invitrogen, Carlsbad, CA, USA) was used for the analysis of 5 mL serum. The manufacturer's protocol was followed for the activation, wash, and desorption steps of the RPC18 beads, with adjustments and optimisations to allow implementation on a 96-channel liquid-handling Hamilton STARplusW pipetting robot. Shortly, RPC18 beads were washed with a 0.1% trifluoroacetic acid solution and peptides were eluted using a 50% acetonitrile solution. The eluates obtained in this last step were transferred into a 96-well plate (*i.e.* eluate plate) and mixed with stabilisation buffer. A portion of these eluates (2 mL) was used for MALDI spotting while the remainder was frozen and stored at -80 °C for later tandem mass spectrometry (MS/MS) experiments. MALDI spotting was performed on the 96-channel liquid handling robot by mixing 2 mL of the sample eluates with 10 mL of an  $\alpha$ -cyano-4-hydroxycinnamic acid solution (0.3 g/L in ethanol/acetone 2:1). Each sample was spotted in quadruplicate onto a MALDI AnchorChip (600 nm; Bruker Daltonics, Bremen, Germany), *i.e.* on the target plate 384 spots were obtained from 96 different samples.

**MALDI-TOF-MS.** MALDI-TOF experiments were performed on an UltraFlex II (Bruker Daltonics) either operating in positive reflectron mode in the  $m/z$  range of 600–4000 or in positive linear mode in the  $m/z$  range of 1000–6500. The spectra were acquired using FlexControl software version 3.0 (Bruker Daltonics). A Smartbeam 200 Hz solid-state laser, set at a frequency of 100 Hz, was used for ionisation. A profile, or summed spectrum, was obtained for each MALDI spot by adding 20 spectra of 60 laser shots each at different rasters. To this end, FlexControl software decided on-the-fly whether or not a scan was used for the summed spectrum. For reflectron mode acquisitions, a resolution higher than 2000 was required. Peaks were detected using the SNAP centroid peak detection algorithm with signal-to-noise (S/N) threshold of 1 and a ‘TopHat’ baseline subtraction. All mass scans not fitting these criteria were excluded. For linear mode acquisitions, a resolution higher than 100 was required and peaks were detected using the centroid peak detection algorithm with S/N threshold of 2 and a ‘TopHat’ baseline subtraction. The measurement of a MALDI spot was finished when 1200 laser shots had been summed in one profile. FlexAnalysis Software 3.0 (Bruker Daltonics) was used for visualisation and data processing.

**Internal calibration of MALDI-TOF serum peptide profiles.** Prior to the measurement of the serum peptide samples, the MALDI-TOF system was externally calibrated using a commercially available peptide mix (Bruker Daltonics). All MALDI-TOF serum peptide profiles were baseline subtracted and internally calibrated using six previously identified peptides (see Results section and Fig. 3a). The TopHat algorithm was used for the baseline subtraction while the Centroid Algorithm, with a S/N threshold of 5 and a height of 80%, was used for the peak detection.

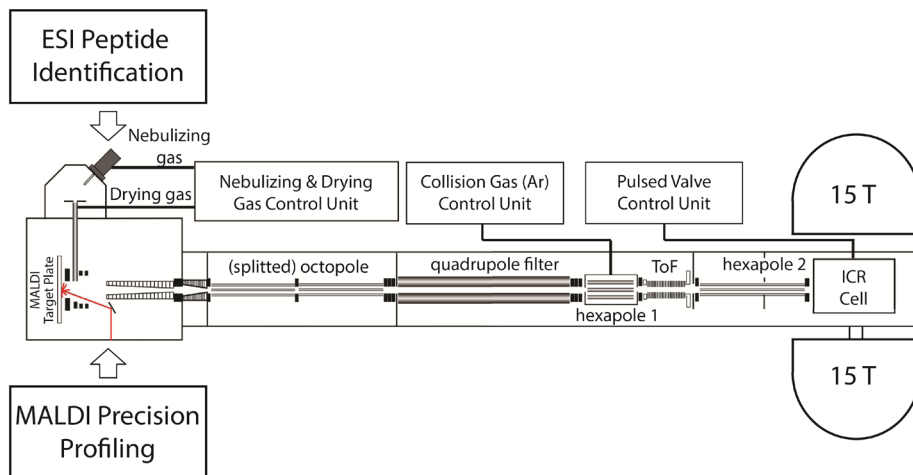
**MALDI-FTICR-MS.** MALDI-FTICR experiments were performed on a Bruker 15 tesla solariX™ FTICR mass spectrometer equipped with a novel CombiSource (Bruker Daltonics, depicted in Fig. 1). The MALDI-FTICR system was controlled by Compass solariXcontrol software and equipped with a Bruker Smartbeam-II™ laser system that operated at a frequency of 500 Hz. The ‘medium’ predefined shot pattern was used for the irradiation. Each mass spectrum was obtained from a single scan of 600 laser shots using 512 K data points. Typically, the target plate offset was 100 V and the deflector plate was

180 V. The ion funnels operated at 100 V and 6.0 V, respectively, with the skimmers at 15 and 5 V. The trapping potentials were set at 0.60 and 0.55 V, the analyzer entrance was maintained at -7 V, and side kick technology was used to further optimise peak shape and signal intensity. The required excitation power was 28% with a pulse time of 20.0 ms. MALDI-FTICR profiles were obtained from the same target plate that had been used for the MALDI-TOF acquisitions. Two acquisition settings were used to obtain profiles in the low-mass and high-mass range, for which only the different parameters are overviewed in Table 1.

**Internal calibration of MALDI-FTICR serum peptide profiles.** DataAnalysis Software 4.0 SP 3 (Bruker Daltonics) was used for the visualisation and the calibration of the spectra. Prior to the measurement of each MALDI plate the FTICR system was externally calibrated using a commercially available peptide mix and a protein mix (Bruker Daltonics). The spectra were then internally calibrated using the same six peptides as in MALDI-TOF profiles (see Results and discussion section, Fig. 3a). Peaks were detected using the FTMS algorithm with a S/N threshold of 5 and using the centroid for peak position with a percentage height of 85. An exclusion mass list was used to filter out all the peaks not included in the evaluation of the methods.

**Direct infusion ESI-FTICR-MS/MS for peptide identification.** The eluates obtained from the RPC18-magnetic bead workup were used for MS/MS experiments (see section ‘Sample collection’). To this end, eluates from 96 different samples were pooled and ultrafiltrated using a 30 kDa filter (Amicon Ultra 0.5 mL; Millipore). The filtrate was concentrated by lyophilisation before RPC18-LC/MS analysis on a splitless nanoLC-Ultra 2D plus system (Eksigent, Dublin, CA, USA) equipped with a PepMap C18 trap column (300  $\mu$ m i.d., 5 mm length; Dionex, Sunnyvale, CA, USA) and a ChromXP C18 analytical column (300  $\mu$ m i.d., 15 cm length; Eksigent). Here, 1 min manual fraction collection was performed at a flow rate of 4 mL/min, with a gradient of 4% to 44% acetonitrile in 0.05% formic acid buffer. Then, the 35 fractions were analysed on the Bruker 15 tesla solariX FTICR mass spectrometer (Fig. 1) controlled by Compass solariXcontrol software after mixing them with 100 mL of spray solution (50:50 MeOH/H<sub>2</sub>O + 0.1% formic acid). Direct infusion ESI experiments were performed at an infusion rate of 2 mL/min. MS/MS

experiments were performed using the quadrupole for precursor ion selection (with an isolation window of 10 mass units), followed by CID in the hexapole.



**Figure 1.** Schematics of the Bruker 15 tesla solariX™ FTICR system equipped with a CombiSource, used for both MALDI precision profiling experiments as well as for direct infusion ESI peptide identification studies.

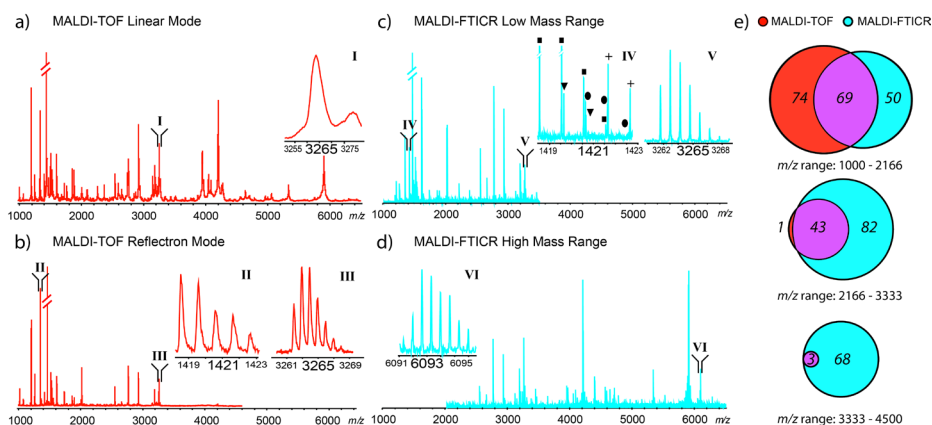
**Table 1.** Instrument parameters that differed between low- and high mass acquisitions using the MALDI solariX™ FTICR system.

Compass solariX control	Low mass acquisition	High mass acquisition
Mass range	$m/z$ 1,013 – 3,500	$m/z$ 2,026 – 6,500
Time of Flight	1.300 ms	3.000 ms
Quadrupole filter Mass	$m/z$ 1,540	$m/z$ 2,500
DC Extract Bias (hexapole)	0.6 V	0.7 V
DC Bias (hexapole 2)	-12 V	-25 V

## RESULTS AND DISCUSSION

In this study, both MALDI-TOF and MALDI-FTICR profiles were obtained from the same serum samples of 96 individuals using fully automated and standardised protocols for collection of human blood samples as well as solid-phase extraction (SPE) with RPC18-functionalised magnetic beads as previously described [17,21,22]. In Fig. 2, typical examples of MALDI peptide profiles are summarised. Figure 2(a) shows a low-resolution linear mode TOF spectrum, as would be common in the early days of peptide profiling [21]. It is well known that in such a profile, peptides are observed as single broad peaks that only allow the determination of an average mass of the peptide. In the corresponding high-resolution reflectron spectrum, the isotopic peaks are resolved and thus measured independently, as is illustrated for the species at  $m/z$  3265 (Fig. 2(b), inset III). Such monoisotopic peptide mass values are advantageous for identification purposes and have become the standard in peptide profiling [22]. Recently, the potential of a high-resolution profile has been exploited in terms of analysing isotope distributions and internal correlations, and applied as a quality control (QC) parameter [17].

In this study, the same set of 96 different serum peptide samples was also measured on a 15 tesla MALDI-FTICR system. These FTICR profiles were obtained using two different acquisition settings in quick succession, *i.e.* for optimal coverage of peptides in the low-mass range as well as in the high-mass range. Typical examples of such ultrahigh resolution profiles are given in Figs. 2(c) and 2(d). All peptides in these spectra were isotopically resolved, with a resolving power 56 000 at  $m/z$  6093 (Fig. 2(d), inset VI). The isotopic resolution of the peptides present in the profile improves the information regarding the chemical composition and allows further QC in clinical studies [17]. On average, the resolving power at  $m/z$  3265 in the MALDI-TOF reflectron mode was 11 000 (Fig. 2(b), inset III), whereas in the corresponding FTICR profiles the resolution of this peptide was 52 000 (Fig. 2(c), inset V). Moreover, the ultrahigh resolving power provided by the MALDI-FTICR system resulted in the detection of overlapping peptides that could not be resolved in reflectron TOF data.



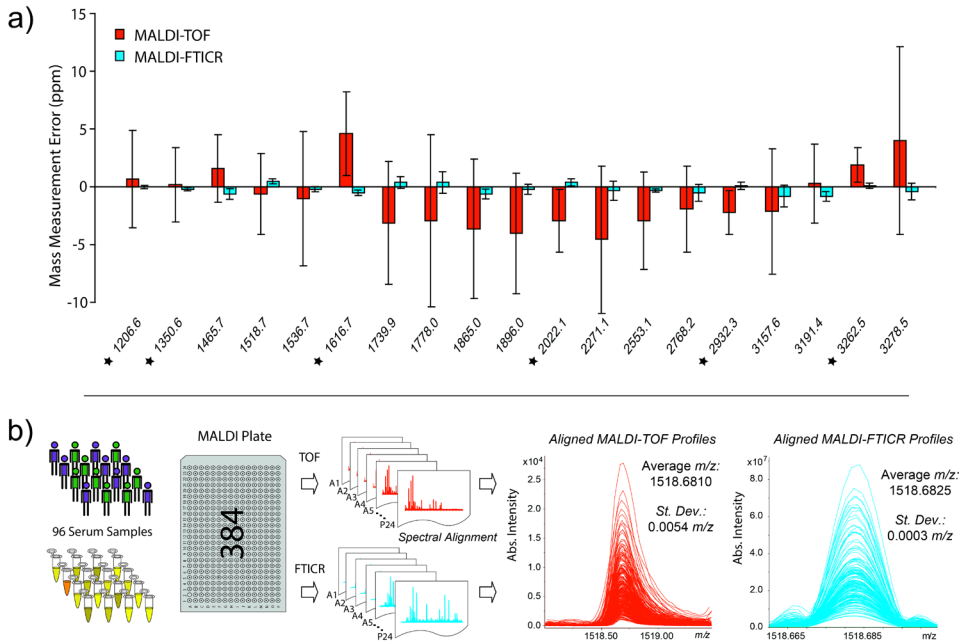
**Figure 2.** MALDI peptide profiles were obtained on both a TOF and an FTICR system, depicted in red and blue, respectively. The peptides were solid-phase extracted from human serum samples. (a) Low-resolution linear mode MALDI-TOF profiles were recorded up to  $m/z$  6500 with all peptides observed as single peaks (inset I). (b) High-resolution reflectron mode MALDI-TOF profiles were recorded up to  $m/z$  4500 with all peptide signals isotopically resolved (insets II and III). (c) A typical example of a low mass range MALDI-FTICR profile. The ultrahigh resolving power allows the detection of four overlapping peptides that could not be resolved in reflectron TOF data (comparing insets II and IV). (d) A typical example of a high mass range ultrahigh resolution MALDI-FTICR profile. (e) Venn diagram depicting the number of observed features in FTICR and reflectron TOF profiles.

This is exemplified for the signal at  $m/z$  1421: the isotopic distribution observed in the TOF profile suggests overlapping, nonresolved species (Fig. 2(b), inset II), which is confirmed in the FTICR spectrum that resolved four different species (Fig. 2(c), inset IV), that can thus be independently quantified. A detailed evaluation on the number of detected peaks, or features, with at least a relative intensity of 0.01% compared to the base peak and a S/N ratio higher than 10 in both TOF and corresponding FTICR profiles resulted in 190 and 315 features, respectively. The results are summarised in Fig. 2(e), in which the features are divided into equidistant parts of the  $m/z$  range of 1000 to 4500. The Venn diagram shows that the number of the features detected in the MALDI-TOF spectra dramatically decreased along the  $m/z$  range, whereas in the corresponding FTICR profiles, the distribution of features over the considered  $m/z$  range was much more homogeneous.



In other words, most peptides in reflectron TOF spectra were profiled at the lower  $m/z$  range and in FTICR profiles all three parts of the  $m/z$  range contained more than 50 features. In order to evaluate the repeatability of peptide mass measurements as well as the corresponding peak intensities in more detail, replicates of an in-house prepared serum sample standard (obtained from a pool of serum samples) were analysed. From this sample 48 MALDI spots were measured in reflectron TOF mode, 48 spots were measured on the new FTICR system using the low-mass acquisition settings, and 48 spots were used to obtain high-mass FTICR profiles. In the TOF profiles 24 peptides were selected for a detailed evaluation, whereas in the low-mass and high-mass FTICR profiles 36 and 21 peptides were used, respectively. The standard deviation of 48 mass measurements at 48 different spots was calculated for each of the considered peptides after internal calibration using known MALDI profile peptides [23]. In the MALDI-TOF profiles this standard deviation varied from 0.70 to 7.7 ppm, with an average of 3.0 ppm. In the low-mass and high-mass MALDI-FTICR profiles this standard deviation varied from 0.19 to 1.0 ppm (average 0.51 ppm) and 0.25 to 2.2 ppm (average 1.1 ppm), respectively. In addition, the repeatability of the intensity measurements was determined from this dataset and results are provided in Supplementary Tables S1, S2 and S3 (see Supporting Information). It was found that the coefficient of variation (CV) was on average 16% in the TOF profiles (varying from 12 to 29% for 24 different  $m/z$ -values), 19% in the low mass FTICR profiles (varying from 9 to 33% for 36 different  $m/z$ -values), and 17% in the high mass FTICR profiles (varying from 11 to 26% for 21 different  $m/z$ -values). From these data it was concluded that the repeatability of the intensity measurement was rather similar for TOF (Ultraflex II) and FTICR (solariX). Aiming for high-throughput screening of clinical samples, a representative set of 96 human serum samples was spotted in quadruplicate on a MALDI target plate after SPE workup. From this plate, 384 MALDI-TOF spectra were obtained as well as 192 corresponding 'low-mass' and 192 'high-mass' FTICR profiles. After a selection procedure based on the presence of calibrant peptides at a predefined S/N level 301 MALDI-TOF and 314 MALDI-FTICR spectra remained [17]. The  $m/z$ -values in the profiles were compared to a previously reported list of 'MALDIMS profile peaks'[23]. Six of these identified peptides were selected for internal calibration, with calculated  $m/z$ -values of monoisotopic peaks of protonated species at 1206.5749 (fibrinopeptide A (FPA) 5-16, P02671), 1350.6284 (FPA

3-16, P02671), 1616.6594 (FPA 1-16 + P, P02671), 2021.1039 (complement C3f fragment, P01024), 2931.2915 (fibrinogen alpha (FGA) chain precursor 576-601, P02671), and 3261.4641 (FGA chain precursor 576-604, P02671). These six peptides with 13 other (unidentified) peptides were selected for further evaluation of their MME and corresponding standard deviation for each individual spectrum. The results were summarised for all selected spectra and are depicted in Fig. 3(a). At a first glance the differences between TOF and FTICR spectra are as predicted, *i.e.* the TOF (red) errors dominate over the FTICR (blue) errors. It can be seen that the average MME of the peptides in 301 TOF profiles varied from -4.5 to 4.6 ppm and in 314 FTICR profiles from -0.8 to 0.9 ppm. Note, however, that some of the MMEs are at a low-ppm level in reflectron mode TOF profiles, and sometimes even lower than in corresponding FTICR profiles. However, the average mass measurement error only informs us about the robustness of the instrument and its calibration. It is the precision of the mass measurement that is crucial for alignment, integration and quantitation. Here the benefit of using FTICR is more obvious, as the standard deviation of the MME for peptides observed in both TOF and FTICR profiles is at least 10-fold better in the ultrahigh resolution spectra. An improved precision is of great value for profiling studies since it simplifies analysis and makes spectral alignment more accurate and robust. Alignment of spectra across all samples and individuals is essential for valid comparisons of profiles in clinical studies. As an example, the results of the standard alignment procedure (see Experimental section) of the peptide at  $m/z$  1518.68 detected in both TOF and FTICR profiles are shown in Fig. 3(b). Note that this alignment was performed on profiles from 96 different individuals and thus included large intensity differences of each peptide up to three orders of magnitude.



**Figure 3. (a) Comparison of mass measurement errors of known peptides in TOF and FTICR profiles. The errors and corresponding standard deviations were determined from replicate measurements of 96 different serum samples. The peptides indicated with an asterisk were used for internal calibration. The peptide identities are as follows: 1206.6=G.EGDFLAEGGGV; 1350.6=D.SGEGDFLAEGGGV; 1465.7=A.DSGEGDFLAEGGGV; 1518.7=ADS(-H2O)GEGDFLAEGGGV; 1536.7=ADSGEGDFLAEGGGV; 1616.7=ADSpGEGDFLAEGGGV; 1739.9=R.NGFKSHALQLNNRQLR; 1778.0=S.SKITHRIHWESASLLR; 1865.0=SSKITHRIHWESASLLR; 1896.0=G.RNGFKSHALQLNNRQLR; 2022.1=SSKITHRIHWESASLLR; 2271.1=S.SRQLGLGPPDVPDHAAYHPF.R; 2553.1=K.SSSYSKQFTSSTSYNRGDSTFES.K; 2768.2=K.SSSYSKQFTSSTSYNRGDSTFESK.Y; 2932.3=K.SSSYSKQFTSSTSYNRGDSTFESKSY.K; 3157.6=R.NVHSGSTFFKYLLQGAIKPPEASFSPR.R; 3191.4=K.SSSYSKQFTSSTSYNRGDSTFESKSYKM.A; 3262.5=K.SSSYSKQFTSSTSYNRGDSTFESKSYKMA.D; 3278.5=K.SSSYSKQFTSSTSYNRGDSTFESKSYKMoxA.D.**

**(b) Comparison of alignment of MALDI-TOF and FTICR profiles, exemplified for the monoisotopic peptide signal at  $m/z$  1518.68. Spectra were obtained from 96 different serum samples, all spotted in quadruplicate on a MALDI AnchorChip target plate. Different colors are applied for the individuals and serum aliquots to indicate heterogeneity in disease state and peptide concentration, respectively.**

Moreover, it should be mentioned that the alignment of FTICR spectra was performed much faster and with fewer (no) errors compared to TOF data. An additional advantage of the improved precision relates to peak identification. When assuming a normal distribution of MMEs in a specific set of profiles, it can be reasoned that the exact mass of the species lies with 95% certainty within two standard deviations from the measured mass. Provided the  $m/z$ -values of unknown peaks are determined at a sub-ppm precision level, the identification window in a database search can also be limited to such a level [19,24]. The MALDI-FTICR precision profiles presented here of human serum peptides contain more than 100 different peptides, all with a measured  $m/z$  value at sub-ppm precision. It is well recognised that a similar mass precision of a large amount of peptides can be obtained from an LC/MS platform with high-end mass analysis, such as the Orbitrap or an ESI-FTICR system [25,26]. Note, however, that an LC/MS approach requires much more measurement time per sample and as a result the compatibility with large cohorts is challenging. The solarix-FTICR system was also used to perform MS/MS experiments. To this end, peptide mixtures were directly infused in ESI mode, isolated in the Q and fragmented using CID gas in the hexapole collision cell (see Fig. 1). These peptide mixtures were obtained after suitable sample cleanup from the magnetic bead eluates (as described in the Experimental section). Five new peptides could be identified from MS/MS experiments, namely EGDFLAEGGGV (from fibrinopeptide alpha chain 5-15, uniprot P02671), AcGEDAAQAEK FQHPG (from serum deprivation-response protein 2-15, O95810), pyroQAGAAGSRMNF R P G V L S (from inter-alpha-trypsin inhibitor H4 650-666, Q14624), AcSETSR T A F G G R R A V P P N N S N A A E D D L P T V E L Q G V V P R (from factor XIIIa 2-60, P00448), and T F G S G E A D C<sub>336</sub>(C<sub>482</sub>) G L R P L F E K K S L E D K T E R E L L E S Y I D G R (from prothrombin 328-363, P00734). Note that the latter peptide contains a cysteinylolation at cysteine 336 (indicated between parentheses) that originates from the disulfide bridge to cysteine 482. All these assignments were carried out with sub-ppm MMEs on both the selected precursor ion as well as the fragment ions (data not shown). Note that the FTICR provides low ppm MMEs (*i.e.* precise data) that allow an accurate overlap between species observed in MALDI profiles and those determined in ESI spectra, such as previously reported with a 2 ppm window [27]. In the current study this window of overlap between ESI and MALDI data was even lower than 1 ppm, *e.g.*

$0.33 \pm 0.48$  ppm for the peptide 650-666 from inter-alpha-trypsin inhibitor H4, and  $-0.34 \pm 0.67$  ppm for the peptide 2-60 from factor XIIIa.

## CONCLUSIONS

In this study, we show that a combination of highly standardised sample workup protocols with ultrahigh resolution FTICR-MS generates serum peptide profiles with very low MMEs. Following the current developments in ultrahigh resolution MS, mass measurement at higher precision improves the alignment of multiple spectra that are needed for comparative analysis. Moreover, MALDI-FTICR spectra provide  $m/z$ -values at sub-ppm precision of all species present in the profiles. From replicate measurements on multiple MALDI spots of an in-house prepared serum sample standard it was concluded that coefficients of variance (CVs) of peak intensities were similar in TOF and FTICR-spectra. A pool of the serum samples was further analysed by direct infusion ESI-FTICR-MS/MS experiments, which resulted in new peptide identifications for five  $m/z$ -values. Ultrahigh resolution peptide profiling builds on recent developments of MS systems and measurement strategies that provide low MMEs that significantly aid data analysis workflows in discovery studies. Such high precision is essential for robust and standardised high-throughput screening of hundreds to thousands of patient samples.

## SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

## ACKNOWLEDGEMENTS

This work is part of the research programme ‘Decrease Colorectal Cancer Death’, which is financially supported by the Center for Translational Molecular Medicine (CTMM)

## REFERENCES

- [1] T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, A. Bairoch, J. J. M. Bergeron. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* 2010, 7, 681.
- [2] J. Schiller, R. Süß, J. Arnhold, B. Fuchs, J. Leßig, M. Müller, M. Petković, H. Spalteholz, O. Zschörnig, K. Arnold. Matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry in lipid and phospholipid research. *Prog. Lipid Res.* 2004, 43, 449.
- [3] R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino, R. D. Hall. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2007, 2, 778.
- [4] R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature* 2003, 422, 198.
- [5] D. van der Merwe, K. Oikonomopoulou, J. Marshall, E. P. Diamandis. Mass spectrometry: uncovering the cancer proteome for diagnostics. *Adv. Cancer Res.* 2006, 96, 23.
- [6] S. Hanash, A. Taguchi. The grand challenge to decipher the cancer proteome. *Nat. Rev. Cancer* 2010, 10, 652.
- [7] M. Palmblad, M. Wetterhall, K. Markides, P. Håkansson, J. Bergquist. Analysis of enzymatically digested proteins and protein mixtures using a 9.4 Tesla Fourier transform ion cyclotron mass spectrometer. *Rapid Commun. Mass Spectrom.* 2000, 14, 1029.
- [8] M. T. Cancilla, S. G. Penn, C. B. Lebrilla. Alkaline degradation of oligosaccharides coupled with matrix-assisted laser desorption/ionization Fourier transform mass spectrometry: a method for sequencing oligosaccharides. *Anal. Chem.* 1998, 70, 663.
- [9] S. Guan, A. G. Marshall, S. E. Scheppele. Resolution and chemical formula identification of aromatic hydrocarbons and aromatic compounds containing sulfur, nitrogen, or oxygen in petroleum distillates and refinery streams. *Anal. Chem.* 1996, 68, 46.
- [10] R. Bischoff, T. M. Luider. Methodological advances in the discovery of protein and peptide disease markers. *J. Chromatogr. B* 2004, 803, 27.
- [11] T. H. Mize, I. J. Amster. Broad-band ion accumulation with an internal source MALDI-FTICR-MS. *Anal. Chem.* 2000, 72, 5886.
- [12] R. D. Smith, G. A. Anderson, M. S. Lipton, L. Pasa-Tolic, Y. Shen, T. P. Conrads, T. D. Veenstra, H. R. Udseth. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2002, 2, 513.
- [13] M. P. Washburn, D. Wolters, J. R. Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, 19, 242.
- [14] V. Kulasingam, M. P. Pavlou, E. P. Diamandis. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer* 2010, 10, 371.
- [15] M. Palmblad, A. Tiss, R. Cramer. Mass spectrometry in clinical proteomics - from the present to the future. *Proteomics Clin. Appl.* 2009, 3, 6.
- [16] A. K. Callesen, J. S. Madsen, W. Vach, T. A. Kruse, O. Mogensen, O. N. Jensen. Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool? *Proteomics* 2009, 9, 1428.
- [17] S. Nicolardi, M. Palmblad, H. Dalebout, M. R. Bladergroen, R. A. E. M. Tollenaar, A. M. Deelder, Y. E. M. van der Burgt. Quality control

based on isotopic distributions for highthroughput MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spectrom.* 2010, 21, 1515.

[18] M. W. Duncan, R. Aebersold, R. M. Caprioli. The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* 2010, 28, 659.

[19] M. Mann, N. L. Kelleher. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. USA* 2008, 105, 18132.

[20] R. Zubarev, M. Mann. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* 2007, 6, 377.

[21] J. Villanueva, K. Lawlor, R. Toledo-Crow, P. Tempst. Automated serum peptide profiling. *Nat. Protoc.* 2006, 1, 880.

[22] C. R. Jimenez, Z. El Filali, J. C. Knol, K. Hoekman, F. A. E. Kruyt, G. Giaccone, A. B. Smit, K.W. Li. Automated serum peptide profiling using novel magnetic C18 beads off-line coupled to MALDI-TOF-MS. *Proteomics Clin. Appl.* 2007, 1, 598.

[23] A. Tiss, C. Smith, U. Menon, I. Jacobs, J. F. Timms, R. Cramer. A well-characterised peak

identification list of MALDI MS profile peaks for human blood serum. *Proteomics* 2010, 10, 3388.

[24] R. A. Zubarev, P. Hakansson, B. Sundqvist. Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Anal. Chem.* 1996, 68, 4060.

[25] J. V. Olsen, J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 2009, 8, 2759.

[26] M. Palmblad, Y. E. M. van der Burgt, E. Mostovenko, H. Dalebout, A. M. Deelder. A novel mass spectrometry cluster for high-throughput quantitative proteomics. *J. Am. Soc. Mass Spectrom.* 2010, 21, 1002.

[27] A. Römpp, L. Dekker, I. Taban, G. Jenster, W. Boogerd, H. Bonfrer, B. Spengler, R. Heeren, P. Sillevius Smit, T. Luider. Identification of leptomeningeal metastasis-related proteins in cerebrospinal fluid of patients with breast cancer by a combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS. *Proteomics* 2007, 7, 474.







# Chapter 3

Quality Control Based on Isotopic Distributions  
for High-Throughput MALDI-TOF and  
MALDI-FTICR Serum Peptide Profiling

*Simone Nicolardi, Magnus Palmblad,  
Hans Dalebout, Marco Bladergroen,  
Rob A. E. M. Tollenaar, André M. Deelder,  
Yuri E. M. van der Burgt*

J. Am. Soc. Mass. Spectrom.  
2010, 21,1515–1525

## **ABSTRACT**

In this study, we have implemented a new quality control (QC) parameter for peptide profiling based on isotopic distributions. This QC parameter is an objective measure and facilitates automatic sorting of large numbers of peptide spectra. Peptides in human serum samples were enriched using reversed-phase C<sub>18</sub>-functionalized magnetic beads using a high-throughput robotic platform. High-resolution MALDI-TOF and ultrahigh resolution MALDI-FTICR mass spectra were obtained and a workflow was developed for automated analysis and evaluation of these profiles. To this end, the isotopic distributions of multiple peptides were quantified from both MALDI-TOF and MALDI-FTICR spectra. Odd peptide isotope distributions in TOF spectra could be rationalized from ultrahigh resolution FTICR spectra that showed overlap of different peptides. The comparison of isotope patterns with estimated polyaveragine distributions was used to calculate a QC value for each single mass spectrum. Sorting these QC values enabled the best MALDI spectrum to be selected from replicate spots. Moreover, using this approach spectra containing high intensities of polymers or other contaminants and lacking peptides of interest can be efficiently removed from a clinical dataset. In general, this method simplifies the exclusion of low quality spectra from further statistical analysis.

## INTRODUCTION

In the past decade, discovery studies in clinical proteomics have evolved synergistically with the technological progress made in mass spectrometry (MS). The growth in the field of MS-based proteomics [1, 2] has been driven by the introduction of new identification techniques [3, 4] and improvements in the resolution, mass accuracy, robustness, and dynamic range of modern mass analyzers [5–8]. From the early days, serum peptide- and protein profiling has played a major role in clinical studies for biomarker discovery, aiming at identifying differences in peptide- and protein levels in serum of healthy and diseased individuals [9–13]. Peptide and protein patterns can change as a result of disease and are thus helpful in both early detection and monitoring the development of the disease. However, serum is a very complex mixture of biomolecules and, therefore, requires appropriate sample workup. In addition to the complexity, the peptide and protein profiles of serum are usually dominated by highly abundant species (the tip of the iceberg) [14]. Different strategies have been followed to reduce the biological complexity of human serum samples and multiple methods and technologies have been developed for the purpose of sample workup [15–19]. The combination of protein and peptide extraction using functionalized magnetic beads with matrix-assisted laser desorption/ionization (MALDI) MS has been widely used for profiling studies [20–22]. One of the advantages of this approach is that automation of both solid-phase extraction (SPE) and MS allows high-throughput screening. Also, for each serum sample, fresh disposable magnetic beads are used, thus avoiding carry-over that may occur when other techniques such as liquid chromatography (LC) are used. Moreover, magnetic beads with a different functionality allow protein- and peptide enrichment based on different chemical-physical interactions, thus broadening the range of components covered. Several studies have focused in detail on sources of variance in profiling experiments and the necessity of standardization in sample collection and preparation has been demonstrated [23–25]. Robotic liquid handling has the advantage of better reproducibility of the SPE and a greatly increased throughput. Robotic platforms enable a fully automated use of magnetic beads and ensure the control of each step in the extraction protocol [26–28]. In addition, the high speed of data acquisition of MALDI MS allows the analysis of thousands of samples in a realistic time period. The application of statistical analysis for the evaluation of the high number of data generated using this strategy, finally enables a more confident identification of

possible biomarkers. In profiling studies, the most commonly used mass analysis in combination with MALDI is time-of-flight mass spectrometry (TOF-MS) [29 –31]. Following the current developments in high-resolution mass spectrometry, MALDI coupled with Fourier transform ion cyclotron resonance (FTICR-MS) has been applied for the same purpose [32, 33]. The high-resolution and low or even sub-ppm mass measurement errors result in more confident peptide and protein assignments. In this study, we compared peptide profiles using the two MS-platforms indicated above, *i.e.*, MALDI-TOF and MALDI-FTICR. With respect to large-scale mass spectrometric proteomic or peptidomic profiling experiments, the quality control (QC) of both the workflow and the data yielded is an arduous task. Variations can occur pre-analytically (blood collection, clotting time, storage conditions), in the analytical workflow, and in the data acquisition (MS experiments). The human eye is extremely good at differentiating multiple pictures, *e.g.*, peptide profiles. However, the visual inspection of all acquired spectra would be time-consuming and subjective. For this reason, various QC procedures and tools have been developed that allow (automated) evaluation of peptide and protein MS profiles [34 –38]. Generally, the signal to noise ratio and/or the threshold of the noise level are taken into account, whereas in MALDI MS protein and peptide profiling the reproducibility of the signal intensity is rigorously evaluated [39,40]. The use of a highly robust protocol including replicate measurements for each sample allows further standardization. The evaluation of noise levels of different datasets (*e.g.*, different MALDI plates) is more cumbersome and prone to bias. In this study, we describe a novel QC approach for the automatic selection of mass spectra, based on peptide isotopic distributions. To this end, the previously used model amino acid averagine was used to carry out detailed comparisons [41]. A recently developed integration method was used to obtain intensity values for each isotope signal within a given distribution [42]. We will show that this QC parameter can clearly discriminate good and poor peptide spectra.

## **MATERIALS AND METHODS**

**Blood Collection and Serum Handling.** Blood samples were collected from patients and healthy volunteers by antecubital venipuncture while the person was seated. Informed consent was obtained from all patients, and the Leiden University Medical Center (LUMC) Medical Ethical Committee approved the studies. All clinical studies were

carried out under supervision of the Department of Surgery at the LUMC. It is outside the scope of this paper to give a detailed description of the various patient cohorts. Subsets of the data will be reported elsewhere. Within 1 h after collection of blood in a 10 mL BD vacutainer tube (containing a clot activator and a gel for serum separation) the sample was centrifuged for 10 min at 1500 x g. The serum was then transferred to a 1 mL polypropylene cryovial, frozen, and stored at -80 °C. Each cryovial was thawed only once and divided over eight 50 µL Matrix 2D barcoded storage tubes (Thermo Fisher Scientific Inc., Hudson, NH, USA) using an 8-channel Hamilton pipetting robot. These tubes were then distributed over eight different 96-sample matrix latch racks, thus yielding eight identical Matrix racks containing 96 different serum samples. These racks were frozen again at -80 °C and a rack was thawed only once for the automated peptide capture procedure. The sample thawing time, *i.e.*, the time between taking a rack out of the freezer and the starting time of the peptide isolation procedure (next section), was between 30 and 60 min for all serum samples. This high degree of standardization in sample handling and sample thawing-time in combination with the use of the 96-channel pipetting robot ensured identical (robust) and simultaneous treatment of all serum samples in the peptide isolation procedure.

#### **High-Throughput Peptide Isolation with RPC18 Dynabeads and MALDI Spotting.**

The isolation of peptides from human serum was performed using reversed-phase C18-functionalized magnetic beads [11, 26, 27]. For each sample, a fresh suspension of paramagnetic monodisperse beads was used and all 96 serum samples in one Matrix rack were incubated with the same lot-number of beads. In this study 10 µL of commercially available RPC18 Dynabeads (Invitrogen, Carlsbad, CA, USA) was used for the analysis of 5 µL of undiluted human serum. The activation, wash, and desorption steps of the RPC18 beads were based on the manufacturer's protocol; however, the protocol was adjusted to enable an optimal implementation on a 96-channel Hamilton STARplus pipetting robot (Hamilton, Bonaduz, Switzerland). In short, a plate containing 96 magnets was designed and built in-house to accommodate a 96-well PCR microtiter plate (MTP) and to allow for optimal settlement of the magnetic beads at the bottom of each well. Furthermore, it was found that additional activation steps were needed, since automated pipetting does not allow the removal of all liquid from the pulled-down magnetic beads.

Thus, for optimal removal of preservatives during the activation step, the RPC18 beads were washed three times with 50  $\mu\text{L}$  of water. For similar reasons, after binding of the peptides to the RPC18 beads three wash steps with 50  $\mu\text{L}$  of a 0.1% trifluoroacetic acid solution were carried out. Finally, the eluates obtained in the peptide desorption step (performed with a 50% acetonitrile solution) were transferred to a 96-well plate and mixed with stabilization buffer. A portion of these eluates (2  $\mu\text{L}$ ) was used for MALDI-spotting while the remainder was frozen and stored at  $-80\text{ }^{\circ}\text{C}$  for future analysis. The 96-channel Hamilton STARplus pipetting robot was used for mixing of sample eluates with MALDI matrix, followed by spotting on a MALDI target plate. To this end, 2  $\mu\text{L}$  of the stabilized eluate was transferred into a fresh 384-well MTP and mixed with 10  $\mu\text{L}$  of  $\alpha$ -cyano-4-hydroxycinnamic acid (0.3 g/L in ethanol:acetone 2:1). Four MALDI spots of each eluate were obtained after quadruplicate spotting on a MALDI AnchorChip (600  $\mu\text{m}$ ; Bruker Daltonics, Bremen, Germany) using 1  $\mu\text{L}$  of the eluate/matrix mixture for each spot.

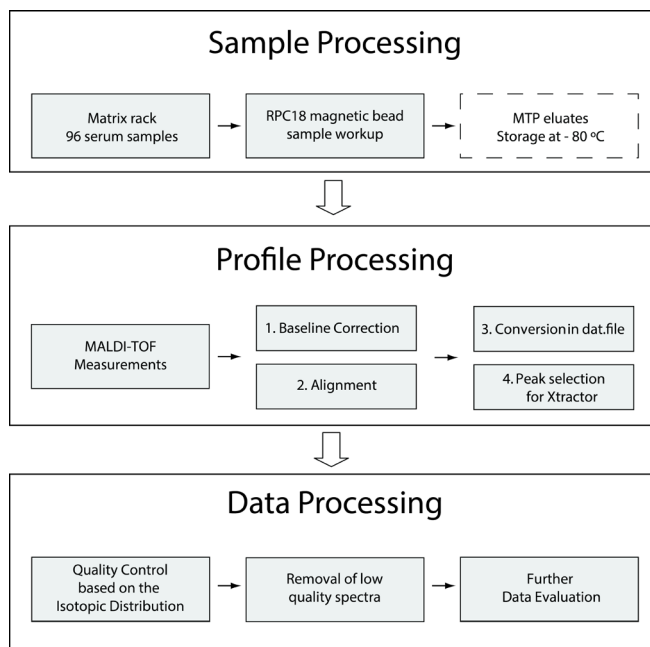
**MALDI-TOF and -TOF/TOF Mass Spectrometry.** After MALDI spotting, the target plate was immediately stored in a storage chamber (RT, 5% oxygen, 95% nitrogen), until transfer into the MALDI-TOF instrument by a CRS F3 robotic arm (Thermo Fisher Scientific Inc., Marietta, OH, USA), as a fully automated plate loader [43]. In this way, all MALDI-TOF measurements were carried out within 12 h after spotting on the 96-channel robot. All MALDI-TOF experiments were performed on an UltraFlex III (Bruker Daltonics) operating in positive reflectron mode in the  $m/z$  range of 600–4000. The spectra were acquired using FlexControl software ver. 3.0 (Bruker Daltonics) with identical data acquisition parameters. A SmartBeam 200 Hz solidstate laser, set at a frequency of 100 Hz, was used for ionization. A profile, or summed spectrum, was obtained for each MALDI-spot by adding 20 spectra of 60 laser shots each at different rasters. FlexControl software decided on-the-fly whether or not a scan was used for the summed spectrum. To this end, a resolution higher than 2000 was required. Peaks were detected using the SNAP centroid peak detection algorithm with signal-to-noise threshold of 1 and a “TopHat” baseline subtraction. All mass scans not fitting these criteria were excluded. The measurement of a MALDI spot was finished when 1200 laser shots had been summed in one profile. The MALDI-TOF spectra were measured from  $m/z$  600 to

4000 and externally calibrated using a commercially available peptide mix (Bruker Daltonics). FlexAnalysis Software 3.0 (Bruker Daltonics) was used for visualization and initial data processing. For identification, the LIFT-TOF/TOF spectra were recorded on the same Bruker Ultraflex III TOF/TOF mass spectrometer. No additional collision gas was applied. Precursor ions were accelerated to 8 kV and selected in a timed ion gate. The fragments were further accelerated by 19 kV in the LIFT cell, and their masses were analyzed after the ion reflector passage.

**Peak Alignment in MALDI-TOF Serum Peptide Profiles.** The workflow for processing peptide profiles is depicted in Figure 1. After external calibration using a commercially available peptide mix, the maximum observed mass shift in the 384 spectra varied from  $\sim 0.1$  Da at  $m/z$  1500 to 0.2 Da at  $m/z$  4000. Before carrying out the alignment a baseline subtraction of all spectra was performed. To allow the alignment of all 384 spectra from one MALDI target plate, at least three peptides at different  $m/z$  values were needed for internal calibration. To compensate for the possible absence of one or two peptides in a spectrum, the following five peptides were selected based on a manual inspection of a few spectra, namely  $m/z$  1465.8, 1778.1, 1865.2, 2602.5, and 2931.5, with a tolerance window of 100 ppm for the  $m/z$  1465.8 peak increasing up to 300 ppm for the highest  $m/z$  value (FlexAnalysis 3.0).

**Determination of Peak Intensities Using Xtractor.** All MALDI-TOF and the MALDI-FTICR spectra were exported as DAT (.dat) and XY (.xy) files, respectively. In both formats (ASCII), all  $m/z$  values with corresponding intensities (*i.e.*, data points) were reported. The simple Xtractor tool was used to determine the intensity of each user-defined peak [42]. To this end, a certain  $m/z$  window was defined for each peak included in a so-called reference file, thus allowing Xtractor to sum all the intensities of the data points within the defined bin. The bin size for all peaks in the MALDI-TOF spectra was fixed at  $\pm 0.49$  Da, the bin size for the peaks in the MALDI-FTICR spectra varied from  $\pm 0.02$  Da at  $m/z$  1465.8 to  $\pm 0.08$  Da at  $m/z$  2931.5. In this way, single intensity values for each isotopic peak within any isotope distribution were obtained. Importantly, Xtractor generates uniform data (peak) arrays regardless of spectra content (Xtractor is an open source tool and can be found at [www.ms-utils.org/Xtractor](http://www.ms-utils.org/Xtractor)).





**Figure 1. Overview of three sequential processing methods for the generation and evaluation of human serum peptide profiles. The novel QC method reported here is carried out in the third step.**

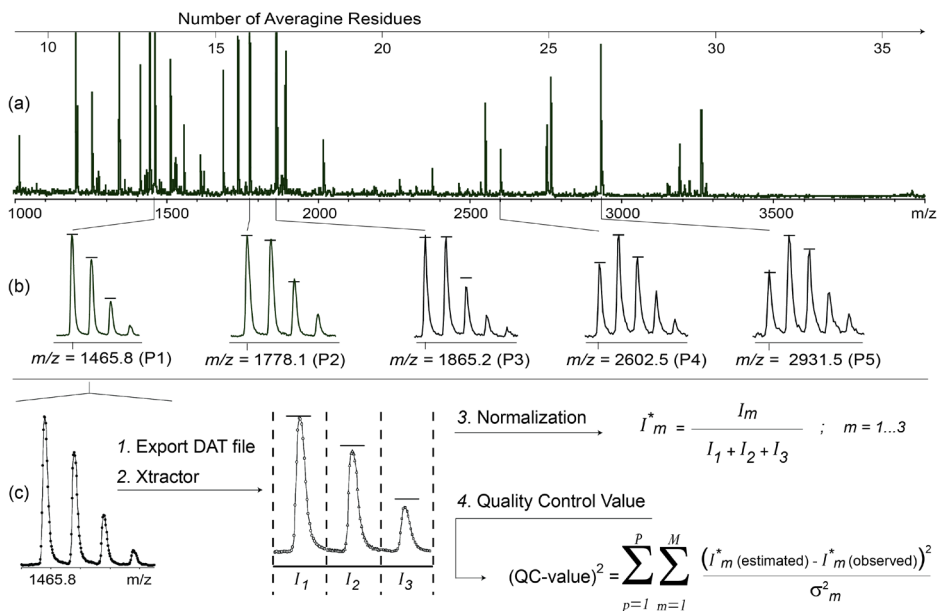
**Calculation of Peak Intensities in Peptide Isotope Distributions.** The observed peptide isotopic distributions were compared with those obtained from the polyaveragine model as explained in Results and Discussion section. The theoretical isotopic distribution of a polyaveragine peptide was calculated using chemcalc ([www.chemcalc.org](http://www.chemcalc.org)). The elemental composition used for the calculation of peak intensities of each individual isotope reflected a singly protonated polyaveragine peptide molecule with integer numbers for all indices of the atoms C, H, N, O, and S. In a similar way, the theoretical isotopic distribution of the peptide at  $m/z$  1465.8 was calculated after identification of this peptide using MALDI-TOF/TOF. Note that each individual isotope within a distribution was considered as one single peak, as described earlier by Rockwood *et al.* [44] Thus, by using the jcamp file from chemcalc, all calculated intensities within the isotopic fine structure of the second and third peak were summed to two individual values only.

**MALDI-FTICR Mass Spectrometry.** All MALDI-FTICR experiments were carried out on a Bruker Daltonics apex-ultra 9.4 tesla FTICR mass spectrometer. This instrument was equipped an Apollo II dual ion source, two ion funnels, a selection quadrupole in the front-end, and an Infinity ICR cell. For the purpose of comparing FTICR-measurements with results obtained from the TOF, manual MALDI-spotting was carried out from a frozen 96-well eluate plate. After thawing the RPC18 eluates at room-temperature, 2  $\mu\text{L}$  of each eluate was transferred into a 384-well plate and mixed with 10  $\mu\text{L}$  of  $\alpha$ -cyano-4-hydroxycinnamic acid (0.3 g/L in ethanol:acetone 2:1). Finally, 1  $\mu\text{L}$  of this mixture was manually spotted on a MALDI Anchor-Chip (600  $\mu\text{m}$ ) in duplicate using a standard Gilson P2 pipette. In this way, 192 spots of 96 different samples were obtained. These spots were measured on the FTICR using a SmartBeam 200 Hz solid-state laser, operated a frequency of 100 Hz. The irradiation spot size was  $\sim 200$   $\mu\text{m}$ . The ions generated from 50 laser shots were accumulated in the hexapole and then transferred through the quadrupole to the collision cell. This process was repeated eight times on different raster spots. The 450 laser shots accumulated in the collision cell were then transferred to the ICR cell for mass analysis. The quadrupole was set to an  $m/z$  of 400 and RF amplitude of 3000  $V_{\text{p-p}}$ . Each spectrum was generated by accumulation of eight scans with 512 K data points. All data were acquired using the Bruker ApexControl software and evaluated using Bruker DataAnalysis software.

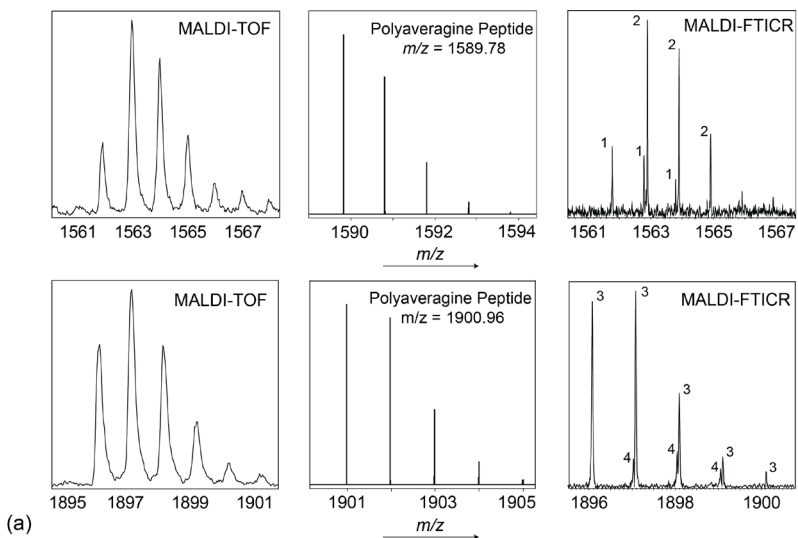
## RESULTS AND DISCUSSION

**Peptide Isotope Distributions in MALDI-TOF Spectra.** The workflow for sample, profile, and data processing is depicted in Figure 1. One of the aims of this study was to demonstrate the use of peptide isotopic distributions of peptides for quality control parameter. For illustration, the isotopic patterns of the five ubiquitous peptides used for alignment are shown in Figure 2. While the selection of these five peptides was an arbitrary process, no software is available that yields a better result without the need of user-defined input parameters. One of these peptides was identified, namely  $m/z$  1465.8, as a fragment of fibrinogen  $\alpha$ -chain (P02671, peptide DSGEGDFLAEGGGVR). Previous reports corroborate this identification [45]. Using the elemental composition of this peptide the theoretical isotopic distribution was calculated (see the Materials and Methods section). It should be emphasized, however, that identification of the peptides is

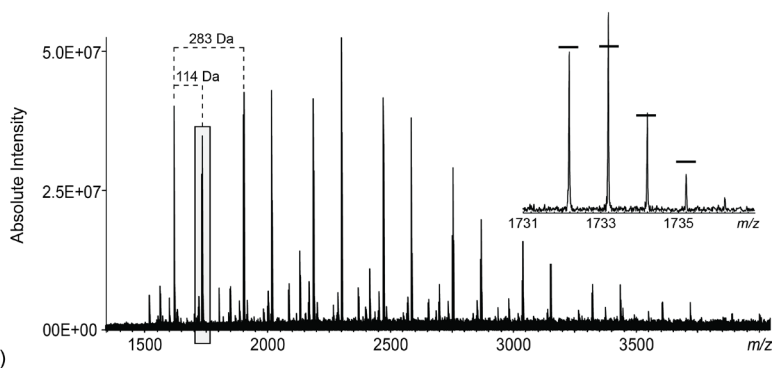
not necessary for the evaluation of isotopic distributions, but clearly becomes important at the stage a biomarker candidate has been detected. Different strategies, followed by adjustments, have been proposed for the estimation of isotopic distributions of unidentified peptides [36, 46, 47]. In this study, we used the “averagine” model, proposed by Senko et al. to construct peptides from a hypothetical average amino acid [41]. This is sufficient for the estimation of isotopic distribution of unknown peptides since we are not searching for the best fit, but only need to discriminate between peptide and nonpeptide (*e.g.*, polymer, matrix cluster) distributions. The estimated values of the five peptides are depicted in Figure 2. These are obtained by first transferring a peptide  $m/z$  value into a neutral mass (*i.e.*, [ $m/z$  value–proton–water]), then dividing this number by the exact mass of averagine (*i.e.*, 111.0543 Da), and multiplying the result with the indices of the molecular formula of averagine (*i.e.*, C<sub>4.9384</sub>H<sub>7.7583</sub>N<sub>1.3577</sub>O<sub>1.4773</sub>S<sub>0.0417</sub>). Finally, one “water-molecule” and a proton are added to create the corresponding “polyaveragine” molecular formula and  $m/z$  value. Due to rounding of the indices, the  $m/z$  values of the observed peptide and the corresponding polyaveragine are not identical. It is clear from Figure 2 that the isotopic patterns of the five peptides used for alignment were in good agreement with the polyaveragine distributions. A quantitative evaluation of these isotope distributions will be given in the next section. Differences with estimated distributions may be indicative of the presence of peptides that contain post-translational modifications (PTMs) or metal cations, or may point at overlapping peptides. Two examples of the latter cause of aberrant isotopic distributions are revealed in Figure 3a. Here, the isotope distributions of two peptides present in some of the MALDI-TOF profiles did not match those of the corresponding polyaveragine peptides. Subsequent high-resolution FTICR mass analysis of the same  $m/z$  values clearly showed overlap of two different peptides. Further analysis showed that the weighted sum of two estimated polyaveragine distributions (with a mass difference of one Da) matched the distributions observed in the TOF spectra. Differences with estimated distributions can be also indicative of sample contamination, such as polymers [48]. An example of MALDI-FTICR spectrum that contains polymer Tinuvin-622 is shown in Figure 3b. The comparison of the isotopic distribution of the polymer at  $m/z = 1732$  with that estimated from a polyaveragine peptide with an  $m/z$  of 1787 clearly shows that the species at  $m/z = 1732$  is not a peptide.



**Figure 2.** (a) MALDI-TOF peptide profile obtained from human serum after sample workup using RPC18-functionalized magnetic beads. (b) Observed isotopic distributions of five peptides chosen for alignment of all serum peptide profiles. The horizontal lines at the top of the peaks represent the calculated values of the corresponding polyaveragine peptides. (c) Profile processing results in a QC value for each individual mass spectrum. First, the intensity of each isotope peak is determined using Xtractor. Then, after normalization, the observed intensities are compared with those estimated from the polyaveragine peptide. For a more accurate determination of the deviation, the random noise variance is taken into account. Finally, the sum of thus obtained deviations for all considered peptides represents the QC value of the MALDI-TOF peptide profile in (a).



(a)



(b)

**Figure 3. (a) Examples of deviations from the polyaveragine model due to overlap of different peptides. Left panel: enlarged part of a MALDI-TOF human serum peptide profile showing odd isotopic distributions. Middle panel: calculated isotope distributions of the corresponding polyaveragine peptides. Right panel: enlarged part of the MALDI-FTICR mass spectrum of the same spot reveals the presence of overlapping peptides 1 and 2 that cause the odd isotopic distribution in the MALDI-TOF spectra, and overlapping peptides 3 and 4. (b) Example of deviation from the polyaveragine model due to different peak identity. A MALDI-FTICR spectrum dominated by polymers of tinuvin-622 (poly-(N- $\beta$ -hydroxyethyl-2,2,6,6-tetramethyl-4-hydroxy-piperidinyl succinate). The enlarged part shows the isotopic distribution of the polymer  $m/z$  1732 and the isotopic distribution calculated from a polyaveragine peptide with  $m/z$  1787 (horizontal bars).**

**Quality Control Based on Peptide Isotope Distributions.** The intensities of the observed peaks in each peptide isotope distribution were determined using Xtractor (see the Materials and Methods section). As an example, for the peptide at  $m/z$  1865.22 the intensity of the monoisotopic was determined as the sum of all measured values at  $m/z$   $1865.22 \pm 0.49$ , the intensity of the second isotope was determined as the sum of all measured values at  $m/z$   $1866.22 \pm 0.49$ , and so forth (Figure 2). The window of 0.49 Da was chosen to ensure total quantification of each isotope. This window of 0.49 Da does not reflect the actual peak width, *i.e.*, the resolution of the peptides in the MALDI-TOF spectra varied from about 7000 at  $m/z$  1465.8 and about 10,000 at  $m/z$  2931.5. It should be noted that the density of data points decreases with increasing  $m/z$  values in a TOF spectrum. As a result, the number of data points within each bin differs between different peptides. However, within one peptide isotopic pattern, the number of data points is the same for each single isotopic peak, at least for the five peptides considered here. Three intensity values were obtained for all five peptides previously discussed in the spectral alignment part using the Xtractor tool. All isotope intensities were then normalized using the sum of the first three intensities within the distribution. These intensities were compared with the corresponding abundances in a polyaveragine isotopic distribution, as previously described by Palmblad et al. [49]. Here, the total signal variance is taken into account in the parameter  $\sigma_m^2$ . In the aforementioned study, carried out using electrospray ionization (ESI) FTICR-MS,  $\sigma_m^2$  contained two terms, namely, experimental signal variance, which was approximately proportional to signal strength, and noise variance, which was constant. In MALDI spectra; however, every single laser shot yields a different amount of ions. As a result, the noise levels are not constant. In general, in the MALDI-TOF experiments performed in this study the noise variance decreased at increasing  $m/z$  values, whereas the opposite was true for the MALDI-FTICR spectra. For the purpose of biomarker discovery and subsequent validation a detailed evaluation of noise variance is pivotal. In this respect, various studies have been reported to determine sources of variance and to provide tools for quantification [14, 25, 31, 34]. The aim of this work is the development of a new tool for evaluation of spectral quality to improve future biomarker discovery studies. To this end, for further calculations the total signal variance  $\sigma_m^2$  was estimated as follows. For one single MALDI target plate the noise levels in TOF- and FTICR profiles were determined for each of the five considered peptides as an

average from 10 different samples (summarized in Table 1). Thus,  $\sigma_m^2$  was considered constant at a specific  $m/z$  value for each sample on one MALDI target plate. Based on visual inspection of the spectrum, an “empty”  $m/z$  region was selected for each peptide used for alignment to calculate a corresponding  $\sigma_m^2$  value. Using these  $\sigma_m^2$  values the deviations between estimated and observed isotopic distributions were determined and summed to one single QC value for each spectrum according to eq 1:

$$(\text{QC-value})^2 = \sum_{p=1}^P \sum_{m=1}^M \frac{(I_m^*(est.) - I_m^*(obs.))^2}{\sigma_m^2} \quad (1)$$

in which P and M are the number of peptides and isotopes considered per peptide, respectively;  $I_m^*(est.)$  and  $I_m^*(obs.)$  are the intensities of the estimated and observed isotopic peaks normalized for the sum of the intensities of the first three isotopic peaks in the corresponding isotopic distributions; the  $\sigma_m^2$  value is the random variance noise calculated in a small set of analyzed spectra.

As an example, the QC values of 384 mass spectra (quadruplicate spots of 96 different serum samples on one MALDI target plate) are depicted in Figure 4. Note that a low QC value corresponds with a high quality spectrum. For most samples with a high quality profile, a similar QC value was observed for the four replicates, as exemplified for S21 in Figure 4. From this it was concluded that MALDI-spotting followed by mass analysis showed good repeatability. QC based on isotopic distribution is a powerful tool for the evaluation of samples that yield low quality spectra. In some cases, only one out of four spectra was poor, as exemplified for S83, in other cases three out of four replicates were poor (e.g, S4 and S49). For this reason it is proposed to select the “best” spectrum of each sample for statistical analysis, *i.e.*, with the lowest QC value. In cases that all four replicates yielded empty spectra (e.g, S58) or showed poor QC values (indicated with circled sample codes in Figure 4) the sample was excluded from further analysis. Applying the selection of one high quality spectrum for each sample ensures standardized further statistics, whereas upon considering all replicate spots the final profile is obtained after averaging of two, three or four spectra.

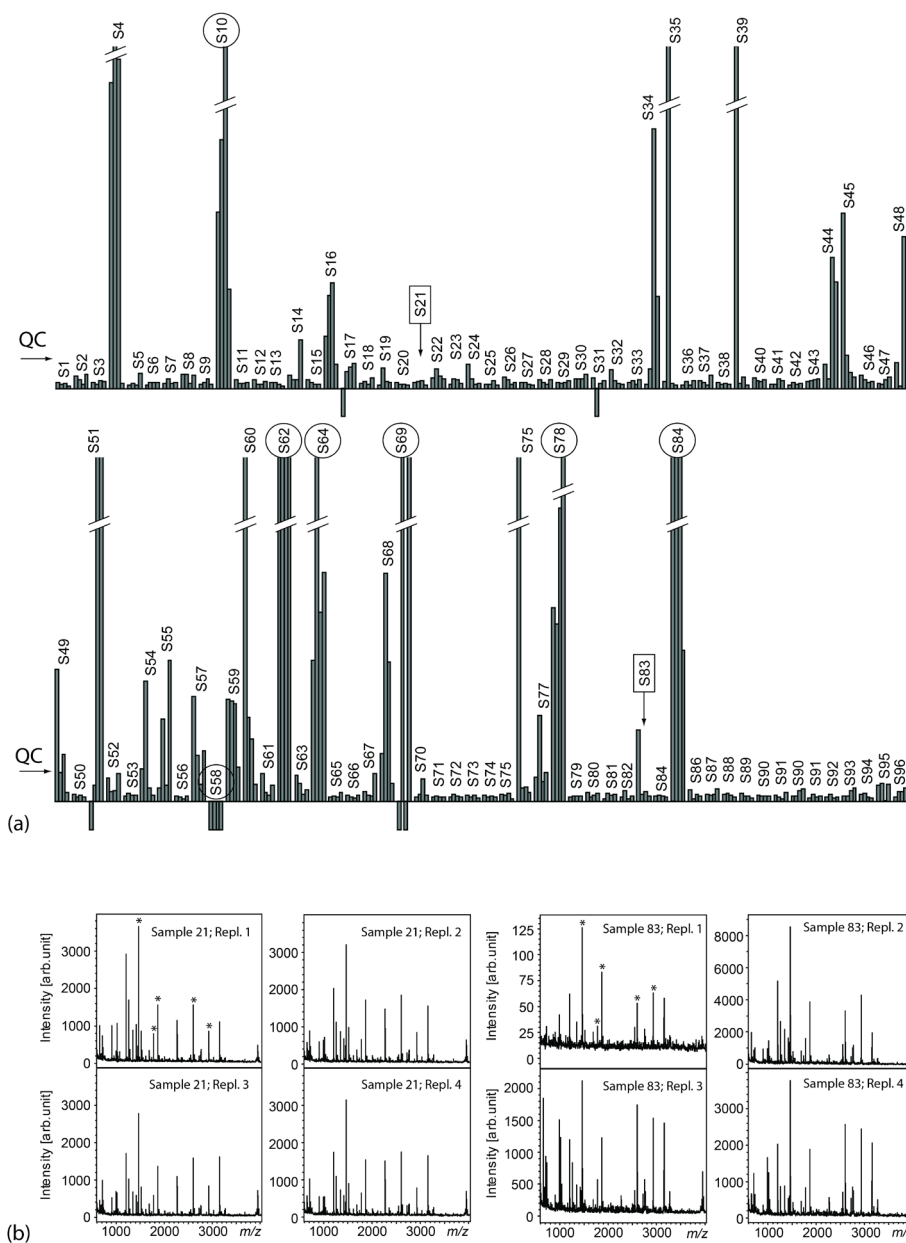
Another way to evaluate the similarity of multiple MALDI profiles is using the dot-product similarity to a reference spectrum as proposed by Yergey and coworkers [35].

In such a comparison, all signal is taken into account, including background peaks. The quality of spectra has been evaluated in detail counting the number of peaks in each replicate measurement [50]. Note that the number of peaks in a spectrum may be high as a result of contaminants and as such is earmarked “good quality”. The same problem will arise upon using total ion intensities, which is exemplified in Figure 5. Here, the FTICR spectrum of no. 3 contains mostly polymers with high signal-to-noise ratios and will result in a false positive. As an alternative, peak picking algorithms are widely used to determine intensities of specific compounds in a spectrum. In combination with peak picking the quality of one single isotope distribution can be determined using the “sophisticated numerical annotation procedure” or SNAP (Bruker). When summarizing the SNAP quality factors of the five selected peptides, a similar trend as in our proposed quality values was observed. However, the extraction of SNAP quality factors from large numbers of mass spectra is time-consuming due to necessary peak picking and deconvolution of each spectrum. Moreover, in the case of a low quality spectrum these processes are often not robust and make automation more difficult. When manually applying SNAP on the same five peptides used for alignment, out of the four replicates in Figure 4, the same profiles were selected as low quality.

**Comparative Analysis of MALDI-TOF and MALDI-FTICR Spectra.** For the purpose of comparing TOF and FTICR serum peptide profiles, the same 192 spots on a MALDI target plate were measured both on a MALDI-TOF and a MALDI-FTICR instrument. The QC value for each TOF spectrum as well as for each FTICR spectrum was determined according to eq 1. The results are plotted in Figure 5 after sorting. Here, it can be observed that a small fraction of the spectra showed a drastically increased QC value that is indicative for a poor spectral quality. It was verified that all TOF-spectra of MALDI-spots with high QC value also yielded low quality FTICR-spectra. In other words, trend analysis of both TOF- and FTICR data resulted in a similar selection of low quality data. It is important to note that the choice of a cut-off point in Figure 5 will be different for each MALDI target plate considered. Any chosen cut-off point for the discrimination between the high and the poor quality spectra based on QC values will introduce false-positive and false-negative results. For a quantitative analysis, the normalized intensities estimated from the polyaveragine peptides, the average ratios of the

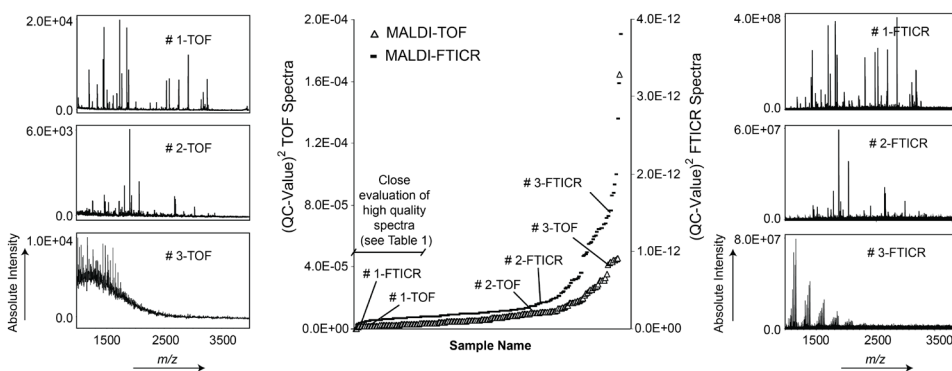


isotopic distributions and their coefficients of variation were evaluated for the five peptides discussed earlier (see Figure 2) and the results are summarized in Table 1. In this evaluation, the 50 MALDI-TOF and the 50 MALDI-FTICR mass spectra with lowest QC values were selected from the results obtained from 192 peptide profiles in Figure 5. It can be seen from Table 1 that the observed isotope ratios were in good agreement with those estimated from the polyaveragine distribution. For example, the averaged isotope ratio of the first and the second peak in the distribution of the peptide at  $m/z$  1865 was 1.1 in the TOF spectra and 1.0 in the FTICR, which is in good agreement with the estimated value of 1.0 from the polyaveragine peptide with  $m/z$  1886. The results in Table 1 show that the coefficients of variation of the first isotope ratios were within 6% and 13% for MALDI-TOF and MALDI-FTICR, respectively. With respect to the possibility of overlap of multiple peptides, it was observed from MALDI-FTICR data that the  $m/z$  regions of the five peptides manually selected for the QC method contained a singly species. Finally, if in the serum, glycopeptides would be present in which the contribution of the glycan part is relatively large, these would be lost upon the RPC18 extraction. This leaves inherent deviations from the polyaveragine peptide as the only explanation for different isotopic distributions in the five peptides. Thus, the observed deviations can be attributed to the differences between the real and the estimated composition.



**Figure 4. (a) Histogram of 384 QC values of 96 human serum samples measured in quadruplicate with MALDI-TOF. The low values (short bars) indicate high quality spectra while a negative value is automatically assigned when an empty spectrum is recorded. Based on these QC values it was found that from the total batch of recorded peptide profiles,**

63 spectra were of poor quality, *i.e.*, 16% of the data were not suitable for further analysis. The cut-off point is indicated with an arrow in the plot. For seven samples, all four replicates yielded a poor quality spectrum (indicated with a circle over the sample code). As a result, when taking into account only the spectrum with the highest quality for each sample, 93% of the data were of high quality and included for further analysis. (b) Two examples of four replicate MALDI-TOF spectra of samples S21 and S83. The calculation of the QC values for each of these peptide profiles allowed the selection of one spectrum for further analysis. Ordering according to the QC method resulted in sample 23 in 4-1-2-3 and in sample 83 in 4-2-3-1. Ordering according to the sum of SNAP quality factors of the same five peptides (indicated with an asterisk in the first spectrum) resulted for samples 23 and 83 in 1-2-4-3 and 2-4-3-1, respectively.



**Figure 5.** Plot of QC values derived from 192 MALDI-TOF and MALDI-FTICR human serum peptide profiles. These 192 profiles were obtained from 96 different samples after duplicate spotting on a MALDI target plate. The QC values are calculated applying eq 1 to the isotopic distributions of five peptides selected in Figure 2. After sorting the MALDI spot numbers by their QC values, clear trends in peptide profile quality are observed.

**Table 1.** The averages of the normalized intensities of the isotope peaks and the coefficients of variation, calculated from the isotopic distribution of five different peptides, are reported. The data were calculated using 50 MALDI-TOF and 50 MALDI-FTICR mass spectra with the lowest QC-values obtained from the measurements of 96 human serum peptide samples (see Figure 4). The  $\sigma_m^2$  is the averaged random noise variance measured in sub-set of MALDI-TOF and MALDI-FTICR in  $m/z$  windows of 5 Da near the monoisotopic peak of each peptide and containing only background signal.

Peptide $m/z$	MALDI-TOF				MALDI-FTICR				Polyaveragine Peptides			
	$I_1^*$ (obs.) (RSD)	$I_2^*$ (obs.) (RSD)	$I_3^*$ (obs.) (RSD)	$\sigma_m^2$	$I_1^*$ (obs.) (RSD)	$I_2^*$ (obs.) (RSD)	$I_3^*$ (obs.) (RSD)	$\sigma_m^2$	Calc. $m/z$	$I_1^*$ (estimated)	$I_2^*$ (estimated)	$I_3^*$ (estimated)
1465.8*	0.44 (2%)	0.36 (2%)	0.19 (7%)	3178	0.49 (4%)	0.36 (4%)	0.15 (10%)	6.4E10	1476.7	0.45	0.37	0.18
1778.1	0.37 (1%)	0.39 (2%)	0.25 (4%)	2757	0.40 (4%)	0.39 (4%)	0.21 (6%)	7.6E10	1786.9	0.39	0.39	0.22
1865.2	0.36 (1%)	0.39 (1%)	0.25 (2%)	2349	0.39 (3%)	0.40 (3%)	0.21 (4%)	8.9E10	1886.9	0.38	0.39	0.24
2602.5	0.29 (4%)	0.38 (3%)	0.33 (2%)	1505	0.30 (6%)	0.41 (5%)	0.29 (6%)	1.3E11	2604.3	0.28	0.40	0.32
2931.5	0.26 (3%)	0.39 (2%)	0.35 (2%)	833	0.26 (7%)	0.40 (5%)	0.34 (5%)	1.4E11	2943.5	0.24	0.40	0.36

\* Peptide identified as a fragment of  $\alpha$ -chain fibrinogen (P02671). The normalized intensities, calculated from the theoretical isotopic distribution, are the following:  $I_1^* = 0.48$ ;  $I_2^* = 0.36$ ;  $I_3^* = 0.16$

## CONCLUSION

In this study, the isotopic distributions of a small set of peptides were used as a quality control parameter for the evaluation of MALDI-TOF and MALDI-FTICR spectra. This novel QC method allows a semi-automated selection of a good spectrum from replicate measurements of one sample and for the removal of low quality spectra from further statistical analysis, which is necessary for large-scale study. The isotopic distributions of five peptides (present in all serum samples) were determined and evaluated in both high-resolution MALDI-TOF and ultrahigh resolution MALDI-FTICR spectra. The selected isotope patterns were compared with polyaveragine distributions to calculate a quality control value for each single mass spectrum. Sorting the obtained QC values allowed the selection of the best MALDI spectrum from replicate spots and the removal of low quality spectra from further data analysis. Examples of odd peptide

isotope distributions in MALDI-TOF spectra could be rationalized from corresponding ultrahigh resolution FTICR profiles that showed overlap of different peptides. As was earlier shown and has been discussed by other groups, multiple factors can invalidate the results of peptide profiling studies. To obtain a confident result, the full workflow of a profiling study should be highly standardized. The novel QC parameter used in this study is objective and facilitates further data analysis in large scale clinical studies. The proposed QC value can be determined for any isotopically-resolved MALDI-profile, *i.e.*, this method is independent of the type of instrument, the manufacturer, or the software used for data processing. The use of this parameter in combination with other quality criteria that take into account signal-to-noise, reproducibility, and baseline threshold should result in a better selection of high-quality spectra. The integration of all such metrics for quality evaluation into one single method is under development.

## **ACKNOWLEDGEMENTS**

The authors acknowledge that this work is part of the research program “Decrease Colorectal Cancer Death,” which is financially supported by the Center for Translational Molecular Medicine (CTMM).

## REFERENCES

1. Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome. *Mol. Cell. Proteom.* 2002, 1, 845–867.
2. Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* 2003, 422, 198–207.
3. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101(26), 9528–9533.
4. Coon, J. J. Collisions or Electrons? Protein Sequence Analysis in the 21<sup>st</sup> Century. *Anal. Chem.* 2009, 81(9), 3208–3215.
5. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.* 2006, 78, 2113–2120.
6. Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer Source. *J. Am. Soc. Mass Spectrom.* 2009, 20, 1391–1396.
7. Schäfer, R. UltrafleXtreme: Redefining MALDI-TOF-TOF Mass Spectrometry Performance. *LC GC Europe* 2009, (Suppl. S), 26–27.
8. Palmblad, M.; Drijfhout, J. W.; Deelder, A. M. High Resolution Mass Spectrometry for Rapid Characterization of Combinatorial Peptide Libraries. *J. Comb. Chem.* 2010, 12, 65–68.
9. Frank, R.; Hargreaves, R. Clinical Biomarkers in Drug Discovery and Development. *Nat. Rev. Drug Discovery.* 2003, 2, 566–580.
10. Palmblad, M.; Tiss, A.; Cramer, R. Mass Spectrometry in Clinical Proteomics—from the Present to the Future. *Proteomics Clin. Appl.* 2009, 3, 6–17.
11. Villanueva, J.; Philip, J.; Entenberg, D.; Chaparro, C. A.; Tanwar, M. K.; Holland, E. C.; Tempst, P. Serum Peptide Profiling by Magnetic Particle-Assisted, Automated Sample Processing and MALDI-TOF Mass Spectrometry. *Anal. Chem.* 2004, 76, 1560–1570.
12. Veenstra, T. D. Global and Targeted Quantitative Proteomics for Biomarker Discovery. *J. Chromatogr. B* 2007, 847, 3–11.
13. Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. Use of Proteomic Patterns in Serum to Identify Ovarian Cancer *Lancet* 2002, 359, 572–577.
14. Callesen, A. K.; Vach, W.; Jørgensen, P. E.; Cold, S.; Mogensen, O.; Kruse, T. A.; Jensen, O. N.; Madsen, J. S. Reproducibility of Mass Spectrometry Based Protein Profiles for Diagnosis of Breast Cancer Across Clinical Studies: A Systematic Review. *J. Proteome Res.* 2008, 7, 1395–1402.
15. Barnea, E.; Sorokin, R.; Ziv, T.; Beer, I.; Admon, A. Evaluation of Prefractionation Methods as a Preparatory Step for Multidimensional Based Chromatography of Serum Proteins. *Proteomics* 2005, 5, 3367–3375.
16. Luque-Garcia, J. L.; Neubert, T. A. Sample Preparation for Serum/Plasma Profiling and Biomarker Identification by Mass Spectrometry. *J. Chromatogr. A* 2007, 1153, 259–276.
17. Faca, V.; Pitteri, S. J.; Newcomb, L.; Glukhova, V.; Phanstiel, D.; Krasnoselsky, A.; Zhang, Q.; Struthers, J.; Wang, H.; Eng, J.; Fitzgibbon, M.; McIntosh, M.; Hanash, S. Contribution of Protein Fractionation to Depth of

- Analysis of the Serum and Plasma Proteomes. *J. Proteome Res.* 2007, 6, 3558–3565.
18. Nissum, M.; Foucher, A. L. Analysis of Human Plasma Proteins: A Focus on Sample Collection and Separation Using Free-Flow Electrophoresis. *Expert Rev. Proteom.* 2008, 5(4), 571–587.
19. Mauri, P.; Petretto, A.; Cuccabita, D.; Basilico, F.; Di Silvestre, D.; Levrieri, I.; Melioli, G.; Fractionation Techniques Improve the Proteomic Analysis of Human Serum. *Current Pharmaceut. Anal.* 2008, 4, 69–77.
20. Baumann, S.; Ceglarek, U.; Fiedler, G. M.; 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(6), 973–980.
21. Jimenez, C. R.; Koel-Simmelink, M.; Pham, T. V.; van der Voort, L.; Teunissen, C. E. Endogenous Peptide Profiling of Cerebrospinal Fluid by MALDI-TOF Mass Spectrometry: Optimization of Magnetic Bead-Based Peptide Capture and Analysis of Preanalytical Variables. *Proteom. Clin. Appl.* 2007, 1, 1385–1392.
22. Alagaratnam, S.; Mertens, B. J. A.; Dalebout, J. C.; Deelder, A. M.; van Ommen, G. B.; den Dunnen, J. T.; 't Hoen, P. A. C. Serum Protein Profiling in Mice: Identification of Factor XIIIa as a Potential Biomarker for Muscular Dystrophy. *Proteomics* 2008, 8, 1552–1563.
23. Rai, A. J.; Vitzthum, F. Effects of Preanalytical Variables on Peptide and Protein Measurements in Human Serum and Plasma: Implications for Clinical Proteomics. *Expert Rev. Proteom.* 2006, 3(4), 409–426.
24. Penno, M. A. S.; Ernst, M.; Hoffmann, P. Optimal Preparation Methods for Automated Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Profiling of Low Molecular Weight Proteins and Peptides. *Rapid Commun. Mass Spectrom.* 2009, 23, 2656–2662.
25. Callesen, A. K.; Madsen, J. S.; Vach, W.; Kruse, T. A.; Mogensen, O.; Jensen, O. N. Serum Protein Profiling by Solid Phase Extraction and Mass Spectrometry: A Future Diagnostics Tool? *Proteomics* 2009, 9, 1428–1441.
26. Villanueva, J.; Lawlor, K.; Toledo-Crow, R.; Tempst, P. Automated Serum Peptide Profiling. *Nature Protoc.* 2006, 1(2), 880–891.
27. Jimenez, C. R.; El Filali, Z.; Knol, J. C.; Hoekman, K.; Kruyt, F. A. E.; Giaccone, G.; Smit, A. B.; Li, K. W. Automated Serum Peptide Profiling Using Novel Magnetic C18 Beads Off-Line Coupled to MALDI-TOFMS. *Proteomics Clin. Appl.* 2007, 1, 598–604.
28. Villanueva, J.; Nazarian, A.; Lawlor, K.; Tempst, P. Monitoring Peptidase Activities in Complex Proteomes by MALDI-TOF Mass Spectrometry. *Nature Protoc.* 2009, 4(8), 1167–1183.
29. Dekker, L. J.; Boogerd, W.; Stockhammer, G.; Dalebout, J. C.; Siccama, I.; Zheng, P.; Bonfrer, J. M.; Verschuuren, J. J.; Jenster, G.; Verbeek, M. M.; Luider, T. M.; Sillevius Smitt, P. A. MALDI-TOF Mass Spectrometry Analysis of Cerebrospinal Fluid Tryptic Peptide Profiles to Diagnose Leptomeningeal Metastases in Patients with Breast Cancer. *Mol. Cell. Proteom.* 2005, 4, 1341–1349.
30. de Noo, M. E.; Mertens, B. J. A.; Özalp, A.; Bladergroen, M. R.; van der Werff, M. P. J.; van de Velde, C. J. H.; Deelder, A. D.; Tollenaar, R.

- A. E. M. Detection of Colorectal Cancer Using MALDI-TOF Serum Protein Profiling. *Eur. J. Cancer* 2006, 42, 1068–1076.
31. de Noo, M. E.; Tollenaar, R. A. E. M.; Özalp, A.; Kuppen, P. J. K.; Bladergroen, M. R.; Eilers, M. R. P.; Deelder, A. M. Reliability of Human Serum Protein Profiles Generated with C8 Magnetic Beads Assisted MALDI-TOF Mass Spectrometry. *Anal. Chem.* 2005, 77, 7232–7241.
32. Stoop, M. P.; Dekker, L. J.; Titulaer, M. K.; Lamers, R. A. N.; Burgers, P. C.; Sillevius Smitt, P. A. E.; van Gool, A. J.; Luider, T. M.; Hintzen, R. Q. Quantitative Matrix-Assisted Laser Desorption Ionization-Fourier Transform Ion Cyclotron Resonance (MALDI-FT-ICR) Peptide Profiling and Identification of Multiple-Sclerosis-Related Proteins. *J. Proteome Res.* 2009, 8, 1404–1414.
33. Römpp, A.; Dekker, L.; Taban, I.; Jenster, G.; Boogerd, W.; Bonfrer, H.; Spengler, B.; Heeren, R.; Sillevius Smitt, P. A.; Luider, T. M. Identification of Leptomeningeal Metastasis-Related Proteins in Cerebrospinal Fluid of Patients with Breast Cancer by a Combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS. *Proteomics* 2007, 7, 474–481.
34. Harrington, P. B.; Vieira, N. E.; Chen, P.; Espinoza, J.; Nien, J. K.; Romero, R.; Yergey, A. L. Proteomic Analysis of Amniotic Fluids Using Analysis of Variance-Principal Component Analysis and Fuzzy Rule-Building Expert Systems Applied to Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Chemomet. Intell. Lab. Syst.* 2006, 82, 283–293.
35. Olson, M. T.; Blank, P. S.; Sackett, D. L.; Yergey, A. L. Evaluating Reproducibility and Similarity of Mass and Intensity Data in Complex Spectra-Applications to Tubulin. *J. Am. Soc. Mass Spectrom.* 2008, 19, 367–374.
36. Cairns, D. A.; Perkins, D. N.; Stanley, A. J.; Thompson, D.; Barrett, J. H.; Selby, P. J.; Banks, R. E. Integrated Multi-Level Quality Control for Proteomic Profiling Studies Using Mass Spectrometry. *BMC Bioinf.* 2008, 9, 519.
37. Bons, J. A. P.; de Boer, D.; van Dieijen-Visser, M. P.; Wodzig, W. K. W. H. Standardization of Calibration and Quality Control Using Surface Enhanced Laser Desorption Ionization-Time of Flight-Mass Spectrometry. *Clin. Chim. Acta.* 2006, 366, 249–256.
38. McLerran, D.; Grizzle, W. E.; Feng, Z.; Bigbee, W. L.; Banez, L. L.; Cazares, L. H.; Chan, D. W.; Diaz, J.; Izbicka, E.; Kagan, J.; Malehorn, D. E.; Malik, G.; Oelschlager, D.; Partin, A.; Randolph, T.; Rosenzweig, N.; Srivastava, S.; Srivastava, S.; Thompson, I. M.; Thornquist, M.; Troyer, D.; Yasui, Y.; Zhang, Z.; Zhu, L.; Semmes, O. J. Analytical Validation of Serum Proteomic Profiling for Diagnosis of Prostate Cancer: Sources of Sample Bias. *Clin. Chem.* 2008, 54(1), 44–52.
39. Albrethsen, J. Reproducibility in Protein Profiling by MALDI-TOF Mass Spectrometry. *Clin. Chem.* 2007, 53(5), 852–858.
40. Callesen, A. K.; DePont Christensen, R.; Madsen, J. S.; Vach, W.; Zapico, E.; Cold, S.; Jørgensen, P. E.; Mogensen, O.; Kruse, T. A.; Jensen, O. N. Reproducibility of Serum Protein Profiling by Systematic Assessment Using Solid-Phase Extraction and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 2008, 22, 291–300.
41. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* 1995, 6, 229–233.



42. Selman, M. H. J.; McDonnell, L. A.; Palmblad, M.; Ruhaak, L. R.; Deelder A. M.; Wuhler, M. Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* 2010, 82, 1073–1081.
43. McDonnell, L.A.; van Remoortere, A.; van Zeijl, R.; Dalebout, H.; Bladergroen, M.R.; Deelder, A.M. Automated Imaging MS: Toward High Throughput Imaging Mass Spectrometry. *J. Proteom.* 2010, 73, 1279–1282.
44. Rockwood, A. L.; Haimi, P. Efficient Calculation of Accurate Masses of Isotopic Peaks. *J. Am. Soc. Mass Spectrom.* 2006, 17, 415–419.
45. Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; Sanchez-Carbayo, M.; Holland, E. C.; Cordon-Cardo, C.; Scher, H. I.; Tempst, P. Differential Exoprotease Activities Confer Tumor-Specific Serum Peptidome Patterns. *J. Clin. Investig.* 2006, 116(1), 271–284.
46. Breen, E. J.; Hopwood, F. G.; Williams, K. L.; Wilkins, M. R. Automatic Poisson Peak Harvesting for High-Throughput Protein Identification. *Electrophoresis* 2000, 21, 2243–2251.
47. Valkenburg, D.; Jansen, I.; Burzykowski, T. A Model-Based Method for the Prediction of the Isotopic Distribution of Peptides. *J. Am. Soc. Mass Spectrom.* 2008, 19, 703–712.
48. Sachon, E.; Matheron, L.; Clodic, G.; Blascoa, T.; Bolbacha, G. MALDI TOF-TOF Characterization of a Light Stabilizer Polymer Contaminant from Polypropylene or Polyethylene Plastic Test Tubes. *J. Mass Spectrom.* 2010, 45, 43–50.
49. Palmblad, M.; Buijs, J.; Håkansson, P. Automatic Analysis of Hydrogen/Deuterium Exchange Mass Spectra of Peptides and Proteins Using Calculations of Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* 2001, 12, 1153–1162.
50. Dekker, L. J.; Dalebout, J. C.; Siccama, I.; Jenster, G.; Sillevius Smitt, P. A.; Luider, T. M. A New Method to Analyze Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Peptide Profiling Mass Spectra. *Rapid Commun. Mass Spectrom.* 2005, 19, 865–870





# Chapter 4

Identification of Human Serum Peptides in  
Fourier Transform Ion Cyclotron Resonance  
Precision Profiles

*Simone Nicolardi, Hans Dalebout,  
Marco R. Bladergroen, Wilma E. Mesker,  
Rob A. E.M. Tollenaar, André M. Deelder,  
and Yuri E. M. van der Burgt*

Int. J. Proteomics.  
2012, 804036

## **ABSTRACT**

The continuous efforts to find new prognostic or diagnostic biomarkers have stimulated the use of mass spectrometry (MS) profiles in a clinical setting. In the early days (about one decade ago), a single low-resolution mass spectrum derived from an individual's body fluid was used for comparative studies. However, a peptide profile of a complex mixture is most informative when recorded on an ultrahigh resolution instrument such as a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer. In this study we show the benefits of the ultrahigh resolving power and the high mass accuracy and precision provided by an FTICR mass spectrometer equipped with a 15-tesla magnet. The ultrahigh-resolution data not only allow assignment of fragment ions with high charge states (4+, 5+) but also enhance confidence of human serum peptide identifications from tandem MS experiments. This is exemplified with collision-induced dissociation (CID) and electron transfer dissociation (ETD) data of middle-down-sized endogenous or protein-breakdown peptides that are of interest in biomarker discovery studies.

## INTRODUCTION

Mass spectrometry- (MS-) based proteomics is a principal platform in systems biology, that is, the integrated approach of different technical disciplines to study the physiological processes in a cell or tissue [1–3]. The diagnosis of a disease or monitoring new medicines in a (pre)clinical setting can be performed either through (theoretically) full proteome analysis or by molecular profiling of large cohorts of biological samples. Nowadays, the first approach is technically feasible, however at low throughput [4]. In the latter high-throughput approach the relevant differences between numerous peptides and proteins are mapped in affected tissues or body fluids of healthy and diseased individuals [5, 6]. To this purpose differential analyses of protein expression patterns in cells, tissue, or body fluids such as plasma or cerebrospinal fluid (CSF) are performed. These patterns can change as a result of disease and are thus helpful in both early detection and monitoring the development of the disease. Additionally, detection of biomarkers can also play a significant role in prevention. Proteomic approaches have been successfully used to obtain information on the state of protein circuits inside tumor cells and at the tumor-host interface [7]. Moreover, the analysis of peripheral blood (serum samples) for biomarkers has emerged as a promising tool for cancer diagnostics, disease monitoring, and prognostication of the patient. Unfortunately, many biomarker discovery profiling studies have not provided sufficient information on the identity of the peptides or proteins of interest. One reason for this incompleteness is the technical limitation of the used MS-platform with respect to resolving power and MS/MS capabilities. In this work we show the benefits of using high-end Fourier transform ion cyclotron resonance (FTICR) MS for identification of human serum peptides. These benefits are even more clear when taking into account that many of these peptides are endogenous or result from protein degradation.

**Ultrahigh Resolution and- Precision Provided by FTICR.** It is well established that the sub-ppm mass accuracy of Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) improves the result and increases the confidence of identifications [8, 9]. Previously, we have shown that a combination of highly standardized sample workup protocols with FTICR precision profiles results in very low mass measurement errors (MMEs) [9]. Moreover, FTICRMS is a powerful tool in tackling sample complexity due

to its ultrahigh mass resolution characteristics [10]. Several recent instrumental and software developments (improved electronics and computer technology) have made FTICR-MS accessible to a broader research community, including the life science area and the clinic. With suitable control and proper calibration, high field FTICR-MS routinely achieves an MME lower than 1 ppm [11, 12]. At very high fields, such as in recently available 15-tesla instruments, MMEs at a ppb level can be reached. The confidence in peptide and protein identifications improves significantly with increasing mass accuracy, especially if mass accuracies below 1 ppm are reached. Moreover, an increased mass accuracy assists in the identifications of peptides that are not in a database (*e.g.*, unsequenced species, posttranslational modifications). The potential of sub-ppm mass accuracy allows for in-depth comparisons of complex peptide mixtures obtained from patient and control body fluids. In order to properly use accurate mass values obtained from modern high-end mass analyzers guidelines have been reported with respect to statistical mass accuracy [13].

**Tandem MS Using Electron Transfer Dissociation.** Peptide sequencing with electron transfer dissociation (ETD) MS/MS has proven to be an extremely powerful and complementary tool next to the widely used collision-induced dissociation (CID) [14, 15]. In a bottom-up proteomics workflow the peptide sequence coverage is often improved with combined CID and ETD fragmentation, either performed on-line in an alternating way or off-line in a sequential experiment [16, 17]. Initially, ETD was implemented on ion trap mass spectrometers and this is still the most commonly applied platform to perform peptide sequencing experiments. In the field of MS-based proteomics it is now well known that ETD is especially favorable for sequencing larger peptides with higher charge states (*i.e.*, 3+, 4+), for middle-down and top-down proteomics, and for determining the location and identity of posttranslational modifications (PTMs) on peptide backbones [15]. Recent developments have enabled ETD experiments on ultrahigh-resolution mass analyzers such as a time of flight (TOF), Orbitrap, and FTICR [15, 18, 19]. In this work we will present both CID and ETD data obtained from an FTICR system from middle-down-sized peptides that are of interest in biomarker discovery studies.

## MATERIALS AND METHODS

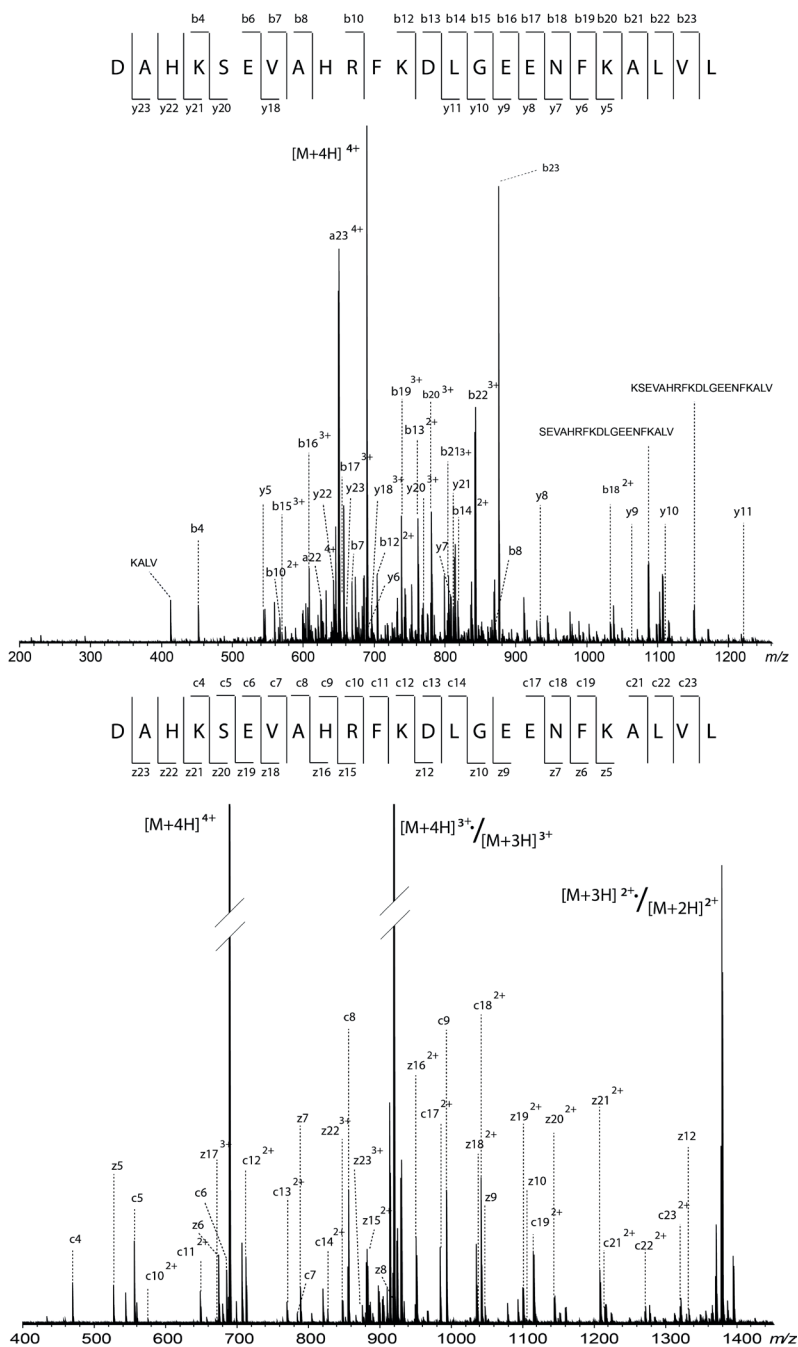
**Sample Collection and Serum Peptide Isolation.** The protocols for both the collection of human blood samples and for serum peptide isolation have been described previously [9]. Informed consent was obtained from all patients and the Leiden University Medical Center (LUMC) Medical Ethical Committee approved the studies. Serum peptides were isolated in a fully automated and standardized fashion using reversed-phase (RP) C18-functionalized magnetic beads. In this study, 10  $\mu\text{L}$  of commercially available RPC18 Dynabeads (Invitrogen, Carlsbad, CA, USA) was used for the analysis of 5  $\mu\text{L}$  serum. The manufacturer's protocol was followed for the activation, wash, and desorption steps of the RPC18 beads, with adjustments and optimizations to allow implementation on a 96-channel liquid handling Hamilton STAR plus pipetting robot [9].

**ESI-FTICR-MS/MS.** The eluates obtained from the RPC18-magnetic bead workup were used for MS/MS experiments (see sample collection). To this end, eluates from 96 different samples were pooled and ultrafiltrated using a 30 kDa filter (Amicon Ultra 0.5 mL; Millipore). The filtrate was concentrated by lyophilization before RPC18-LCMS analysis on a splitless nanoLC-Ultra 2D plus system (Eksigent, Dublin, CA, USA) equipped with a PepMap C18 trap column (300  $\mu\text{m}$  internal diameter, 5 mm length; Dionex, Sunnyvale, CA USA) and a ChromXP C18-analytical column (300  $\mu\text{m}$  internal diameter, 15 cm length; Eksigent). Here, one minute manual fraction collection was performed at a flow rate of 4  $\mu\text{L}$  per minute, with a gradient of 4% to 44% acetonitril in 0.05% formic acid buffer. Then, the 35 fractions were analysed using a Bruker 15 tesla solariX FTICR mass spectrometer equipped with a CombiSource, a quadrupole for precursor ion selection, a hexapole collision cell for CID, and an ETD source on a splitted octopole after mixing them with 100  $\mu\text{L}$  of spray solution (50/50 MeOH/H<sub>2</sub>O 0.1% formic acid) [9]. Direct infusion electrospray ionisation (ESI) experiments were carried out at an infusion rate of 2 microliter per minute. The ion funnels were operated at 100 V and 6.0V, respectively, with the skimmers at 15V and 5V. The trapping potentials were set at 0.60V and 0.55V, the analyzer entrance was maintained at -7V, and side kick technology was used to further optimize peak shape and signal intensity. The required excitation power was 28% with a pulse time of 20.0  $\mu\text{s}$ . MS/MS experiments were performed with the Q (quadrupole) at an isolation window of 10 mass units, followed by

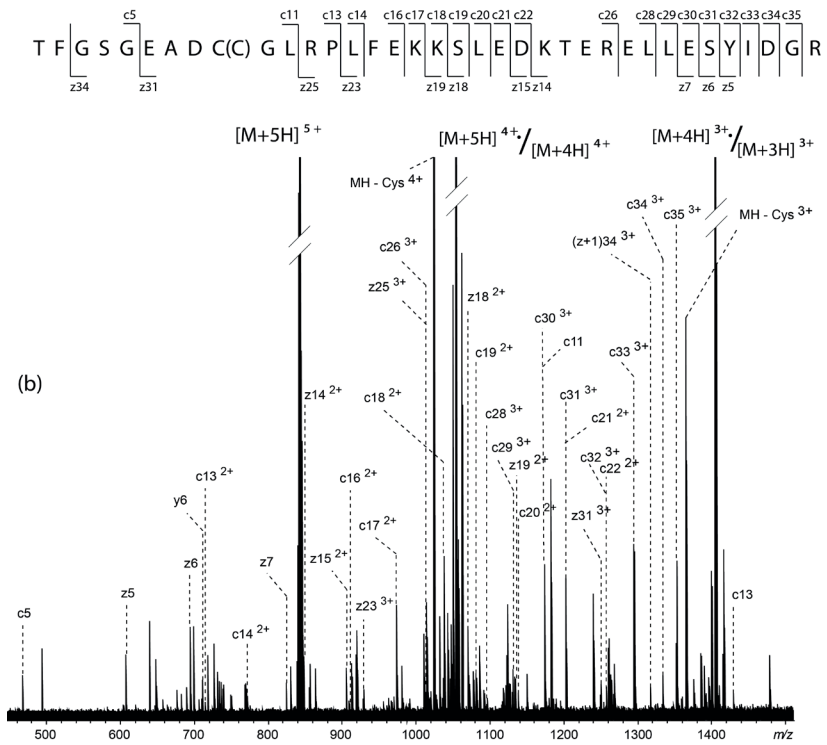
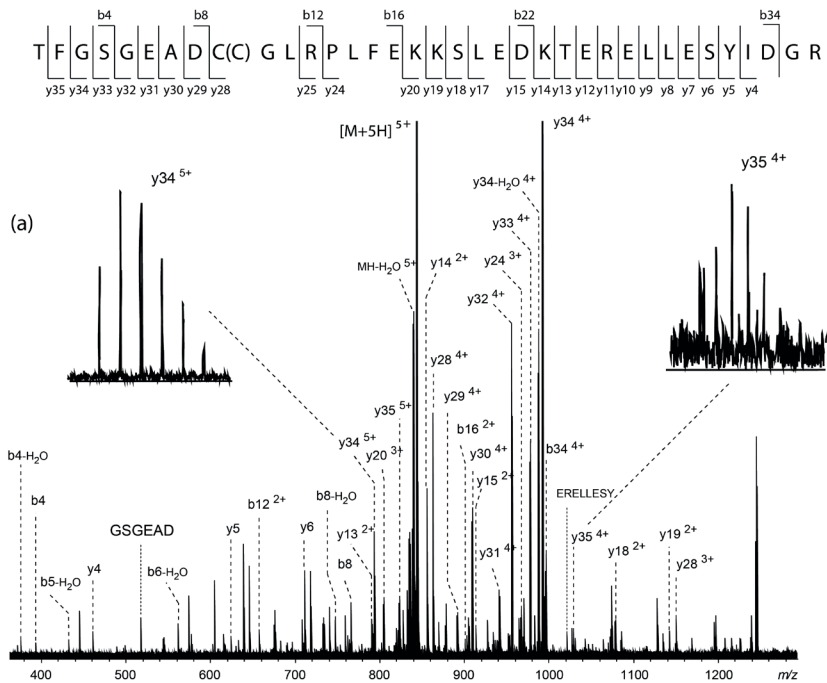


either CID or ETD and fragment ion mass analysis in the ICR cell. For CID experiments, both the collision energy and the accumulation time in the hexapole collision cell were optimized for each precursor ion. Collision energies varied from 5 to 33V while the accumulation times varied from 1 to 5 seconds. For ETD experiments the ETD reagent accumulation time was fixed at 400 ms (fluoranthene from negative chemical ionization (NCI) source). The ETD spectrum was optimized by varying the reaction time (from 10 to 100 ms, with an optimum at 50 ms), varying the peptide (analyte) accumulation time (depending on the precursor ion intensity), and modifying the mirror RF amplitude (280–300 Vpp).

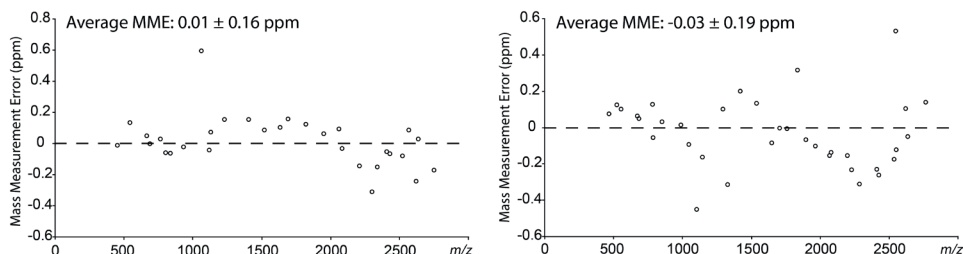
**Data Analysis.** It should be noted that the peptides that were sequenced in this work (obtained from human serum samples) were either endogenous peptides or breakdown products from serum proteins. This implies that any amino acid can be present at the C- and the N-terminus. As a consequence, the interpretation of MS/MS spectra was partly performed manually (de novo) and partly using software from Bruker Daltonics (Data Analysis, BioTools). With a manually determined part of the sequence (sequence tag) a BLAST search was performed. These results were further used to match other parts of the MS/MS data.



**Figure 1.** ESI-FTICR-MS/MS spectra of an albumin peptide isolated from a human serum sample, with sequence coverage in CID (a) and ETD (b).



**Figure 2.** ESI-FTICR-MS/MS spectra of a prothrombin peptide isolated from a human serum sample, with sequence coverage in CID (a) and ETD (b). In order to exemplify the ultrahigh resolving power of the system the insets show the isotopic patterns of fragment ions at  $m/z$  793 [y34(5+)] and at  $m/z$  1028 [y35(4+)].



**Figure 3.** Summary of mass measurement errors of fragment ions in ESI-FTICR-MS/MS CID (a) and ETD (b) spectra of a prothrombin peptide isolated from a human serum sample.

## RESULTS AND DISCUSSION

Matrix-assisted laser desorption ionization (MALDI) precision profiles of serum eluates were obtained using a state-of-the-art solariX-FTICR system equipped with a 15-tesla magnet [9]. Since the purpose of this study is to report peptide identifications, the discussion of such precision profiles is outside the scope of this work. Nevertheless, it should be emphasized that the accurate mass list of peptides obtained from precision profiles allowed a window of overlap between species observed in MALDI profiles and those determined in ESI spectra lower than 1 ppm. For a complex mixture such as serum peptides this can only be achieved in a single-step analysis (*i.e.*, without the need for LC-separation) using the ultrahigh resolving power of FTICR.

The solariX-FTICR system was further used to perform MS/MS experiments aiming for peptide sequencing. To this end, peptide mixtures were directly infused in ESI mode, isolated in the Q, and fragmented through either CID or ETD. These peptide mixtures were obtained after suitable sample clean-up from the magnetic bead eluates and fractionation (as described in Section 2). As an example, the CID and ETD spectra of an albumin fragment (UniProt P02768, number 25–48) are shown in Figure 1. The ultrahigh resolving power of the FTICR allowed confident assignment of many fragment ions, resulting in a peptide sequence coverage of 87% (20 out of 23 peptide bonds) and 91%

from CID and ETD, respectively. From Figure 1 it follows that in the albumin peptide three bond cleavages were not observed upon CID, namely, between the amino acids S-E, H-R, and FK. In the ETD spectrum these bond cleavages, among others, were observed, stressing the complementarity of these two fragmentation methods. The complementarity is also valid vice versa; namely, bond cleavages between amino acids E-E and K-A were observed in CID and not in ETD. Furthermore, it is interesting to report on the signal intensities of the CID and ETD fragments of the peptide with MW 2754 Da: these varied from  $4 \times 10^6$  to  $4 \times 10^8$ , with a precursor ion intensity of  $5 \times 10^8$ , and from  $1 \times 10^7$  to  $3 \times 10^8$ , with a precursor ion intensity of  $3 \times 10^9$ , respectively. This exemplifies the large dynamic range of the FTICR; that is, also the low abundant fragment ions are clearly detectable in an MS/MS spectrum. The second example of complementarity of CID and ETD fragmentation is presented in Figure 2, although not up to a total sequence coverage of 100%. The sequence coverage of the prothrombin peptide was 77% from the CID spectrum and 60% from the ETD-spectrum, whereas the combination of data resulted in 89%. This peptide was identified as a fragment from prothrombin (UniProt P00734, number 328–363), namely, TFGSGEADC(C)GLRPLFEKKSLEDKTERELLE SYIDGR. Note that this peptide contains a cysteinylolation at cysteine 336 (indicated between parentheses) that originates from the disulfide bridge to cysteine 482. Both the ultrahigh resolving power and the subppm MMEs of all fragment ions provided by the FTICR were essential for confident assignments. In the case of the CID spectrum especially, the charge states of most fragment ions obtained from the large precursor peptide with five protons could not have been resolved with relatively low-resolution instrumentation such as an ion trap mass spectrometer. This is exemplified by the insets in Figure 2, showing the isotopic patterns of fragment ions at  $m/z$  793 [y34(5+)] and at  $m/z$  1028 [y35(4+)]. The MMEs are overviewed in Figure 3. The average MMEs for the CID and ETD experiments of the peptide with MW 4210 Da were  $0.01 \pm 0.16$  ppm and  $-0.03 \pm 0.19$  ppm, respectively. Again, the benefits of the large dynamic range (more than two orders of magnitude) of detecting fragment ions become clear from the following number: the signal intensities of the CID fragments of the peptide with MW 4210 Da varied from  $2.5 \times 10^6$  to  $5 \times 10^8$ , with a precursor ion intensity of  $2 \times 10^8$ , and the signal intensities of the ETD fragments of this peptide varied from  $1 \times 10^7$  to  $6 \times 10^8$ , with a precursor ion intensity of  $8 \times 10^9$ .

## REFERENCES

- [1] R. Aebersold and M. Mann, “Mass spectrometry-based proteomics”, *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.
- [2] N. L. Anderson and N. G. Anderson, “The human plasma proteome: history, character, and diagnostic prospects”, *Molecular & Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, 2002.
- [3] T. Ideker, V. Thorsson, J. A. Ranish et al., “Integrated genomic and proteomic analyses of a systematically perturbed metabolic network”, *Science*, vol. 292, no. 5518, pp. 929–934, 2001.
- [4] T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, A. Bairoch, and J. J.M. Bergeron, “Mass spectrometry in high-throughput proteomics: ready for the big time”, *Nature Methods*, vol. 7, no. 9, pp. 681–685, 2010.
- [5] R. Frank and R. Hargreaves, “Clinical biomarkers in drug discovery and development”, *Nature Reviews Drug Discovery*, vol. 2, no. 7, pp. 566–580, 2003.
- [6] S. Surinova, R. Schiess, R. Hüttenhain, F. Cerciello, B. Wollscheid, and R. Aebersold, “On the development of plasma protein biomarkers”, *Journal of Proteome Research*, vol. 10, no. 1, pp. 5–16, 2011.
- [7] D. A. N. Mustafa, P. C. Burgers, L. J. Dekker et al., “Identification of glioma neovascularization-related proteins by using MALDI-FTMS and Nano-LC fractionation to microdissected tumor vessels”, *Molecular and Cellular Proteomics*, vol. 6, no. 7, pp. 1147–1157, 2007.
- [8] M. Palmblad, Y. E. van der Burgt, E. Mostovenko, H. Dalebout, and A. M. Deelder, “A novel mass spectrometry cluster for high-throughput quantitative proteomics.”, *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 6, pp. 1002–1011, 2010.
- [9] S. Nicolardi, M. Palmblad, and P. J. Hensbergen, “Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry”, *Rapid Communication in Mass Spectrometry*, vol. 25, pp. 3457–3463, 2011.
- [10] A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, “Fourier transform ion cyclotron resonance mass spectrometry: a primer”, *Mass Spectrometry Reviews*, vol. 17, no. 1, pp. 1–35, 1998.
- [11] B. Bogdanov and R. D. Smith, “Proteomics by first mass spectrometry: TOP down and bottom up”, *Mass Spectrometry Reviews*, vol. 24, no. 2, pp. 168–200, 2005.
- [12] R. D. Smith, G. A. Anderson, M. S. Lipton et al., “An accurate mass tag strategy for quantitative and high-throughput proteome measurements”, *Proteomics*, vol. 2, pp. 513–523, 2002.
- [13] R. Zubarev and M. Mann, “On the proper use of mass accuracy in proteomics”, *Molecular and Cellular Proteomics*, vol. 6, no. 3, pp. 377–381, 2007.
- [14] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt, “Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9528–9533, 2004.
- [15] C. K. Frese, A. F. M. Altelaar, M. L. Hennrich et al., “Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-orbitrap velos”, *Journal of Proteome Research*, vol. 10, no. 5, pp. 2377–2388, 2011.

- [16] H. Molina, R. Matthiesen, K. Kandasamy, and A. Pandey, "Comprehensive comparison of collision induced dissociation and electron transfer dissociation", *Analytical Chemistry*, vol. 80, no. 13, pp. 4825–4835, 2008.
- [17] D. L. Swaney, G. C. McAlister, and J. J. Coon, "Decision tree driven tandem mass spectrometry for shotgun proteomics", *Nature Methods*, vol. 5, no. 11, pp. 959–964, 2008.
- [18] D. A. Kaplan, R. Hartmer, J. P. Speir et al., "Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer", *Rapid Communications in Mass Spectrometry*, vol. 22, no. 3, pp. 271–278, 2008.
- [19] Y. O. Tsybin, L. Fornelli, C. Stoermer et al., "Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry", *Analytical Chemistry*, vol. 83, pp. 8919–8927, 2011.







# Chapter 5

Mapping O-glycosylation of apolipoprotein C-III  
in MALDI-FT-ICR protein profiles

*Simone Nicolardi, Yuri E. M. van der Burgt,  
Manfred Wuhrer and André M. Deelder*

PROTEOMICS

2013, 13, 992-1001

## **ABSTRACT**

Ultrahigh resolution MALDI-FT-ICR profiles were obtained from human serum samples that were processed using a fully automated RPC18-based magnetic bead method. Proteins were profiled from  $m/z$  value 6630 with a resolving power of 73000 up to  $m/z$  value 12600 with a resolving power of 37000. In this study, a detailed evaluation was performed of the isoforms of apolipoprotein C-III, *i.e.* the different mucin-type core 1 O-glycans with the addition of one or two sialic acid residues. The MALDI-FT-ICR profiles are discussed with regard to reproducibility of the signal intensities as well as the accurate mass measurements. ESI-FTICR-MS/MS analyses of the same serum samples were performed to confirm the identity of apolipoprotein C-III glycoforms.

## INTRODUCTION

Human serum peptide and protein profiling by MALDI coupled with TOF-MS is a well-established discovery strategy for disease-related biomarkers [1]. Studies reported in the previous decade have resulted in a large number of candidate peptides and proteins that were found to correlate with the presence and/or a stage of a certain disease. However, most of these discoveries have not been translated into a diagnostic clinical assay [2–5]. The obvious explanation for this setback is the inherent lack of “depth” in profiling studies, *i.e.* of the existing ten orders of magnitude of concentration range in protein abundance, only the first three or four layers are mapped [5]. Another important reason for the hitherto disappointing outcome of MS-based biomarker studies relates to validation, as robustness of the analytical platform at high-throughput use is essential. It has been shown that the reproducibility of MALDI profiling methods improves using highly standardized sample collection and work-up protocols [6–9]. The use of a fully automated liquid-handling robotic platform results in reproducible SPE peptide purification steps as well as in robust MALDI spotting procedures; thus, standardizing the entire serum profiling workflow. Sample purification and/or separation procedures are needed prior to analysis by MS, because body fluids such as serum are highly complex mixtures. With each additional procedure that is included in the workflow, more effort is required to achieve good robustness of the analytical protocols. While MALDI profiling protocols include off-line peptide or protein purification methods [10], the application of LC together with ESI allows separation of peptide and proteins with on-line MS measurement. The separation efficiency of LC has led to both a larger number of profiled peptides and proteins and to a more quantitative analysis when triple quadrupole (Q) MS was used [11,12]. However, relatively low throughput is inherent to most LCMS methods, which makes the analysis of large sample cohorts difficult and thus impedes the process of biomarker validation. High-throughput approaches pose a large demand on LC equipment. Obviously, chromatographic runs can be shortened (to a certain extent) to alleviate the problem of running into weeks or even months of analysis time. Another way to tackle sample complexity involves the use of ultrahigh resolution instruments such as FT-ICR or Orbitrap R MS. These provide the highest mass measurement precision and accuracy and allow separation of peptides and proteins that would overlap in lower resolution mass spectrometers [13–15]. We recently reported serum peptide profiles ( $m/z$  range from 1000

to 6500 Da) with low parts per million (ppm) mass measurement errors (MMEs) obtained from an MALDI-FT-ICR system equipped with a 15 tesla magnet [16]. It was shown that the sub-ppm mass measurement precision improved the alignment of multiple spectra that are needed for comparative analysis and thus allowed a robust and standardized high-throughput screening of hundreds to thousands of patient samples. In almost any biomarker discovery study larger peptides and proteins are profiled by low resolution MALDI-TOF-MS. Unfortunately, because of the lower resolving power at higher masses as well as the intrinsic more complex isotopic distribution, such larger peptides and proteins are detected as broad peaks. In these cases, the identification and quantification of single species is very challenging with respect to robustness of the method. Moreover, when protein signals in profiles are isotopically resolved the mass differences between corresponding variants can be determined with great accuracy, thus allowing the determination of a point mutation or a PTM [17]. In the current study, we have used a new MALDI-FT-ICR method for profiling of large serum peptides and proteins in the mass range from 6000 to 15 300 Da with high sensitivity and high resolution. Based on the previously reported identifications in MALDI profiles, we focused on apolipoproteins that were observed in the current profiles, namely apolipoprotein C-I (Apo-CI), C-II (Apo-CII), and C-III (Apo-CIII), including its glycosylated isoforms [1, 18, 19]. Apolipoproteins C (Apo-Cs) are present in blood associated, at different degrees, with various lipoproteins particles [20–22]. Several studies have been carried out in order to elucidate the role of Apo-Cs in the metabolism of lipoproteins and results have been comprehensively reviewed [23–25]. Briefly, Apo-Cs are involved in the metabolism of lipoproteins through the activation or inhibition of several triacylglycerol lipases and interaction with different lipoprotein receptors [26]. It is known that Apo-Cs levels are abnormal with certain diseases. For example, individuals affected by diabetes as well as those subjects with hypertriglyceridemia have increased concentration of Apo- CIIIs [27,28]. The risk of coronary heart disease has been also correlated with Apo-CIIIs levels in blood [29, 30]. Cohen et al. recently reported on an MALDI-TOF profiling study showing that Apo-CI and Apo-CIII extracted on C8-coated beads were significantly lower in serum samples of stomach cancer patients as compared to healthy individuals [31]. In addition, the glycosylation pattern of Apo-CIII is being used as a marker for certain congenital disorders of glycosylation (CDGs). In fact, Wopereis et al. found that Apo-CIII

IEF can be used to detect abnormalities in the biosynthesis of core 1 mucin-type O-glycans and reported that patients with CDG type IIe and II f as well as part of patients with an unspecific CDG type II had hypoglycosylated Apo-CIII isoform profiles [32, 33]. Density gradient ultracentrifugation procedures have been applied for the separation of the lipoprotein particles containing Apo-Cs [34]. Other techniques such as delipidation, 2DE [35–37], IEF [38], and LC [39, 40] have been used to further purify and separate Apo-Cs. In the first approach, long centrifugation times are often required, thus decreasing the throughput of the analysis. Apo-Cs have been successfully extracted and profiled using a combination of SPE and MALDI-TOF MS [18, 19, 41–43]. In the present study, we perform magnetic bead (MB) based SPE of serum samples followed by the acquisition of ultrahigh resolution MALDI-FT-ICR profiles including Apo-CI, Apo-CII, and glycosylated Apo-CIII isoforms.

## MATERIALS AND METHODS

**Sample collection.** Human blood samples were collected as previously described [44]. Briefly, blood samples were collected in 10-mL BD Vacutainer tubes (containing a clot activator and a gel for serum separation) from healthy volunteers by antecubital venipuncture and centrifuged for 1 h. Serum was then transferred to a 1-mL cryovial and stored at  $-80\text{ }^{\circ}\text{C}$  until further aliquoting. To this end, 96 cryovials were thawed and serum was distributed over eight racks using an eight-channel liquid-handling robot (Hamilton, Bonaduz, Switzerland). Each rack was stored again at  $-80\text{ }^{\circ}\text{C}$  and thawed only for the automated peptide isolation procedure.

**Serum protein isolation.** Serum samples were anonymized before purification of proteins by SPE either based on magnetic beads (MBs) or on cartridges. Previously, it was shown that these two SPE protocols provided similar results for peptide isolation from serum [45]. In the cartridge-based protocol, serum samples were thawed at room temperature for 1 h, diluted four times with 0.1% acetic acid and loaded (100  $\mu\text{L}$ ) on preconditioned C18-cartridges (SPARK). Each cartridge was washed with 2 mL of 1% acetic acid/2% acetonitrile (ACN) solution and peptides/proteins were eluted with 100  $\mu\text{L}$  of a 50% ACN/0.1% acetic acid solution. This procedure was implemented on a customized SPARK Symbiosis SPE system, specifically adapted to increase sample throughput. The

MB-based protocol was performed for the purification of 96 serum samples from 96 different individuals using a 96-channel liquid handling robot as described previously [16]. After both types of SPE methods, MALDI spotting was performed on the same liquid handling robot by mixing 2  $\mu\text{L}$  of the sample eluates with 10  $\mu\text{L}$  of an  $\alpha$ -cyano-4-hydroxycinnamic acid solution (0.3 g/L in ethanol/acetone 2:1). Each sample was spotted either in duplicate (MBseluates) or quadruplicate (cartridge eluates) onto an MALDI AnchorChip (600  $\mu\text{m}$ ; Bruker Daltonics, Bremen, Germany). Cartridge-based SPE was used for method optimization of the MALDI-FT-ICR experiments (Fig. 1), whereas MBs were used to evaluate the biological variation of the Apo-Cs isoforms that were stored at  $-80\text{ }^{\circ}\text{C}$  for 6 months and thawed at room temperature prior to MALDI spotting (Fig. 3). For identification purposes, proteins were purified from serum samples by RPC4 SPE. Briefly, 300  $\mu\text{L}$  of serum was diluted with 600  $\mu\text{L}$  of a 0.05% formic acid (FA) solution, vortexed and kept at room temperature for 15 min for standardization reasons. The diluted serum was then loaded on a 1 mL BioSelect SPE C4 cartridge (Grace, Deerfield, Illinois, USA) that was prewashed two times with 900  $\mu\text{L}$  of a 95% ACN/4.95% water/ 0.05% FA solution and preconditioned with 900  $\mu\text{L}$  of a 0.05% FA solution. The cartridge was then washed three times with 900  $\mu\text{L}$  of a 0.05% FA solution and the adsorbed peptides and proteins were stepwise eluted with 900  $\mu\text{L}$  of 5, 10, 15, and 20% ACN solution. These five fractions were analysed by MALDI-FT-ICR-MS to determine which fractions contained the relevant apolipoproteins (data not shown). To this end, 1  $\mu\text{L}$  of each of the RPC4 cartridge eluates was mixed with 15  $\mu\text{L}$   $\alpha$ -cyano-4-hydroxycinnamic acid solution (0.3 g/L in ethanol/acetone 2:1) and 1  $\mu\text{L}$  of the mixture was spotted on an MALDI AnchorChip. Direct infusion ESI-FT-ICR analysis was performed only on the RPC4 cartridge fractions that contained the Apo-CIII isoforms. Prior to ESI-FT-ICR MS, 50  $\mu\text{L}$  of each suitable eluate was diluted with 50  $\mu\text{L}$  of a 95% ACN/4.95% water/0.05% FA solution and analyzed by direct infusion.

**MALDI- and ESI-FT-ICR MS.** Both MALDI- and ESI-FT-ICR experiments were performed on a Bruker 15 tesla solariX<sup>TM</sup> FT-ICR mass spectrometer. The MALDI-FT-ICR system was controlled by Compass solariX control software and equipped with a Bruker Smartbeam-II<sup>TM</sup> Laser System that operated at a frequency of 1000 Hz. The “medium” predefined shot pattern was used for spectrum acquisition. Each mass

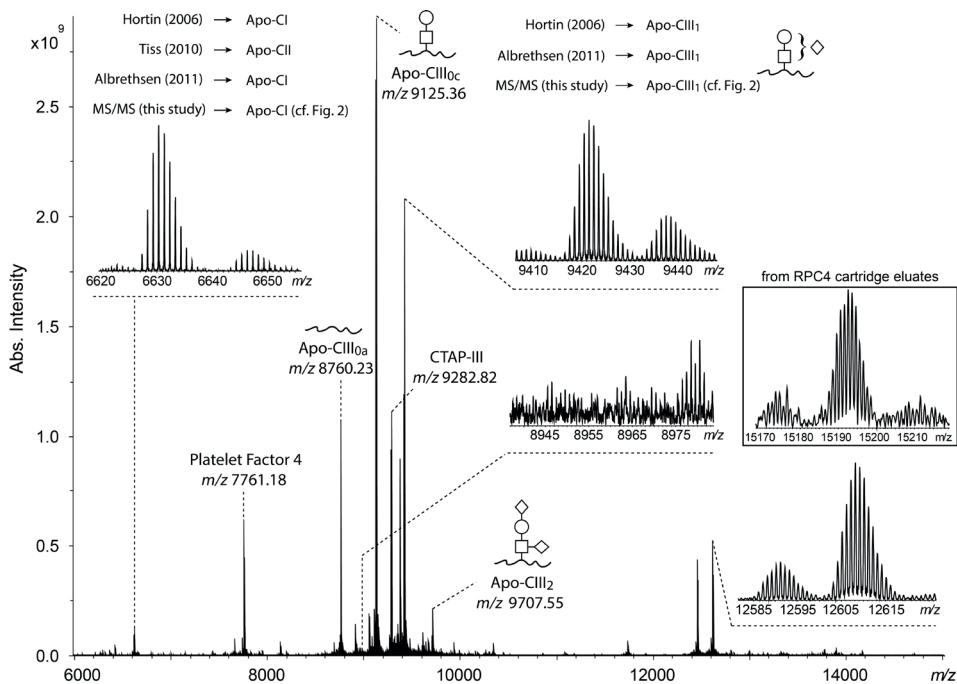
spectrum was acquired from  $m/z$  values 6000 to 15000 and obtained from the average of ten scans of 150 laser shots each using 256 K data points. The random walk option was allowed on a diameter of 300  $\mu\text{m}$ . Typically, the target plate offset was 100 V with the deflector plate set at 180 V. The ions were trapped and measured in the ICR cell using the Gated Trapping option. Both the front and the back trapping potentials were set to 1 V while the front and the back detection voltages were set to 0.4 and 0.45 V, respectively. The ramp time was 0.01 s. The required excitation power was 45% with a pulse time of 15.0  $\mu\text{s}$ . DataAnalysis Software 4.0 SP 3 (Bruker Daltonics) was used for the visualization and the calibration of the spectra while the IsotopePattern software (Bruker Daltonics) was used for the simulation of the isotopic distribution of the Apo-CIII isoforms. Direct infusion ESI-FT-ICR experiments were performed using the same instrumental settings described previously [46]. Briefly, a Q was used for precursor ion selection and a hexapole collision cell for CID. Direct infusion ESI experiments were carried out at an infusion rate of 2  $\mu\text{L}/\text{min}$ . The ion funnels operated at 100 and 6.0 V, respectively, with the skimmers at 15 and 5 V. The trapping potentials were set at 0.60 and 0.55 V, the analyzer entrance was maintained at  $-7$  V, and side kick technology was used to further optimize peak shape and signal intensity. The required excitation power was 28% with a pulse time of 20  $\mu\text{s}$ . MS/MS experiments were performed with the Q at an isolation window of 10 mass units followed by CID and fragment ion mass analysis in the ICR cell. For CID experiments, both the collision energy and the accumulation time in the hexapole collision cell were optimized for each precursor ion. Collision energies varied from 5 to 33 V while the accumulation times varied from 1 to 5 s.

## RESULTS AND DISCUSSION

**MALDI-FT-ICR serum profiles.** A stringent protocol was used for the collection of human blood samples to allow for a valid evaluation of peptide and protein signals in MALDI profiles, and for later comparison of different patient cohorts in a high-throughput manner. All serum samples were processed through a fully automated and standardized workup procedure applying SPE with RPC18 functionalized MBs as described previously [45]. Then, for screening purposes, 1  $\mu\text{L}$  of each serum eluate (*i.e.* obtained after SPE) was measured on an MALDI-TOF system and the remainder of the eluates was stored at  $-80$  °C (data not shown). In this study, these serum eluates were



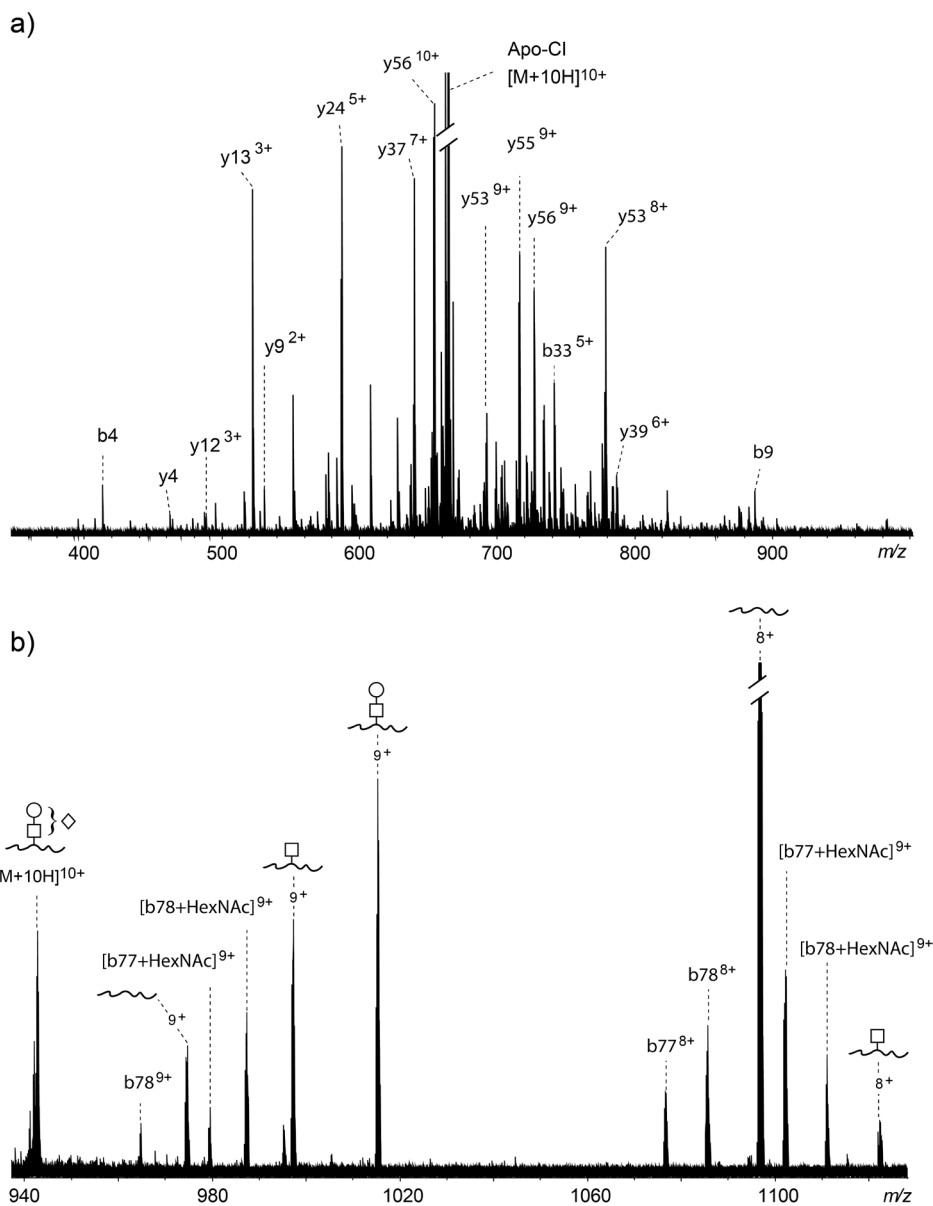
further used to obtain profiles of apolipoproteins. To this end, the eluates were thawed and anonymized, and then spotted onto an MALDI target plate after premixing with matrix solution using the same 96-channel liquid handling platform. MALDI-FT-ICR spectra (profiles) were obtained through an automated acquisition procedure on the solariX™ system (see Section 2). A typical example of an ultrahigh resolution MALDI-FT-ICR serum protein profile recorded from  $m/z$  values 6000 to 15000 is shown in Fig. 1. Note that all species (*i.e.* proteins) in the profile in Fig. 1 are isotopically resolved, with a resolving power of 73000, 51000, and 37000 at  $m/z$  values 6630, 9420, and 12607, respectively. Furthermore, a resolving power of 29000 for the signal at  $m/z$  15193 was determined in the MALDI-FT-ICR protein profile obtained from an RPC4 cartridge eluate, as shown in the inset in Fig. 1 (note that this signal was not observed in the RPC18 MALDI-FTICR protein profile). Recently, we have reported on ultrahigh resolution MALDI-FT-ICR serum peptide profiles up to  $m/z$  value 6500 from similar samples with identical workup. It was shown that MMEs of multiple MS acquisitions were at the sub-ppm level both for individual samples (repeatability) and for hundreds of different serum samples after alignment. Hence, these FT-ICR spectra were established as precision peptide profiles. In the current profile, the  $m/z$  range has been extended up to  $m/z$  value 15000. From Fig. 1, it is evident that the use of an FT-ICR-system allows accurate mass analysis of small proteins in MALDI protein profiles because all species are fully resolved up to an  $m/z$  value of 12615 (or even at  $m/z$  value 15193, Fig. 1 inset). This was not possible on previous MALDI equipment (often TOF mass analyzers).



**Figure 1.** Ultrahigh resolution 15T MALDI-FT-ICR protein profile obtained from human serum purified by RPC18 cartridge-based SPE. Proteins were profiled from  $m/z$  value 6630 with a resolving power of 73000 up to  $m/z$  value 12600 with a resolving power of 37000. The inset shows an enlargement of an MALDI-FT-ICR protein profile obtained from an RPC4 cartridge eluate. The resolving power for the protein at 15193  $m/z$  was 29000. This signal was not observed in the RPC18 MALDI-FT-ICR protein profile. The identification was performed by comparing the observed  $m/z$  values to previously reported lists of (proteolytically degraded) proteins (or polypeptides) (Table 1). The most abundant peaks were identified as O-glycosylated isoforms of apolipoprotein-CIII. ESI-FT-ICR MS/MS experiments confirmed the identity of both Apo-CI and Apo-CIII<sub>1</sub> (Fig. 2).

Moreover, possible peak overlap can be determined in ultrahigh resolution profiles, and peak intensities at high  $m/z$  values can be measured with greater precision [16]. Obviously, the more accurate mass differences between various species within one profile increase confidence levels of their identification (as will be further discussed at the end of Section 3.2, *i.e.* Fig. 3).

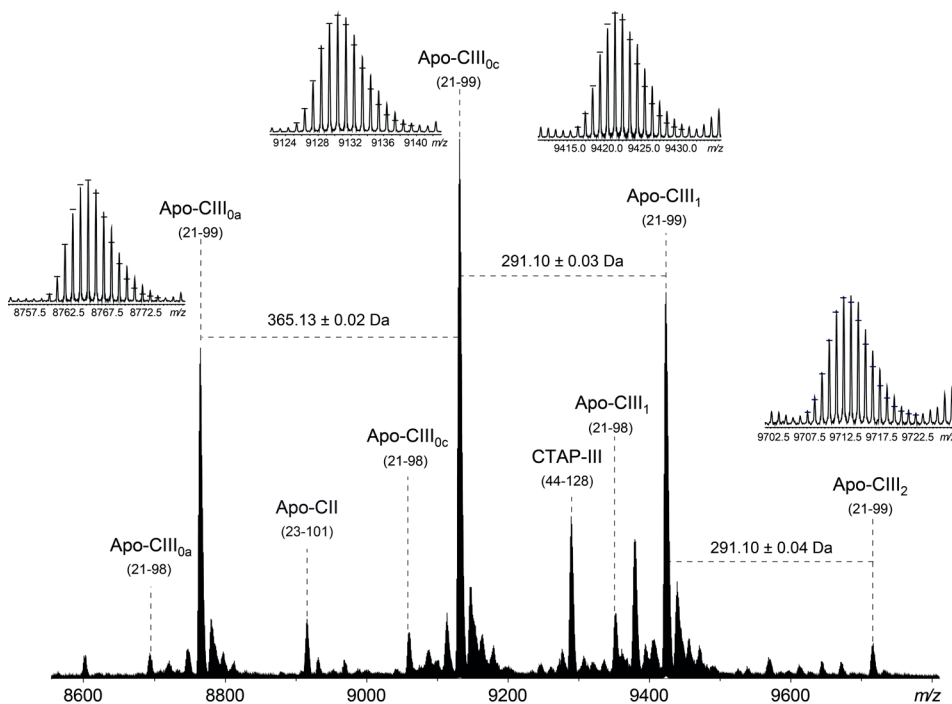
**Apolipoprotein identifications.** Initially, all  $m/z$  values of species observed in the ultrahigh resolution high-mass MALDI-FT-ICR profile (Fig. 1) were compared to previously reported lists of (proteolytically degraded) proteins (or polypeptides) [1, 18, 19]. From these lists it became clear that several apolipoproteins were present in the profile, such as the monoisotopic peak of apo-CI (calculated  $m/z$  value 6627.5131), the monoisotopic peak of apo-CIII (calculated  $m/z$  value 8689.1879), and the monoisotopic peak of apo-CII (calculated  $m/z$  value 8910.3856). A summary of all species in Fig. 1 that could be assigned based on known proteins is given in Table 1. The monoisotopic peak values in our data were in low ppm agreement with the previously reported O-glycosylated isoforms of Apo-CIII, with the remark that glycan linkages cannot be determined from the current experiments. These previously reported Apo-CIII isoforms consist of a single mucin-type core-1 disaccharide galactose linked to N-acetylgalactosamine ( $\text{Gal}\beta 1-3\text{GalNAc}$ ), as such, or include a terminal modification with one ( $\text{NeuAc}\alpha 2-3\text{Gal}\beta 1-3\text{GalNAc}$  or  $\text{Gal}\beta 1-3[\text{NeuAc}\alpha 2-6]\text{GalNAc}$ ) or two ( $\text{NeuAc}\alpha 2-3\text{Gal}\beta 1-3[\text{NeuAc}\alpha 2-6]\text{GalNAc}$ ) sialic acid residues (or N-acetylneuraminic acid (NeuAc)). These three different isoforms of Apo-CIII are commonly referred to as Apo-CIII<sub>0</sub>, Apo-CIII<sub>1</sub>, and Apo-CIII<sub>2</sub>, respectively, the index pointing at the number of sialic acid residues bound to the protein as determined by IEF [32]. Note that in such an IEF profile the Apo-CIII<sub>0</sub> band can include three different isoforms, namely nonglycosylated Apo-CIII, Apo-CIII with one GalNAc monosaccharide, and Apo-CIII with the GalNAc-Gal disaccharide. Since all these three species lack any sialic acids (and are thus referred to as Apo-CIII<sub>0</sub>), it has been proposed to add an index a, b, or c [43]. In this study we did not find any evidence for the presence of Apo-CIII<sub>0b</sub> (calculated  $m/z$  value 8963.3044) in any of the acquired MALDI-FT-ICR profiles of all measured serum samples (Fig. 1). Interestingly, a conflict in assignment of the  $m/z$  value 6627.5 in Fig. 1 was noted, which was typed as Apo-CII by Tiss et al., whereas Hortin and Albrethsen reported Apo-CI. From our ESI-FT-ICRMS/MS data it was concluded that this species (in MALDI observed at  $m/z$  value 6627.5, in ESI selected as precursor ion  $[\text{M}+10\text{H}]^{10+}$ ) matched the Apo-CI amino acid sequence (part of the CID-spectrum is depicted in Fig. 2A).



**Figure 2.** Enlarged parts of the CID spectra of Apo-CI (A) and Apo-CIII<sub>1</sub> (B) obtained with direct infusion ESI-FT-ICR-MS/MS of serum samples that were purified by RPC4 cartridge-based SPE. In both cases, the  $[M+10H]^{10+}$  precursor ion was selected in the quadrupole (Q) to perform MS/MS experiments. In the MS/MS spectrum of Apo-CIII<sub>1</sub> (B) all fragment ions lack the O-glycan, unless indicated explicitly.

**Table 1. List of identified proteins observed in RPC18 magnetic bead and RPC18 cartridge MALDI-FTICR serum profiles in the *m/z*-range of 6,000-15,000. Twelve out of 15 identified proteins belong to the apolipoprotein family including O-glycosylated isoforms of apolipoprotein-CIII. The identity of the two proteins written in bold, namely ApoCI (27-83) and ApoCIII<sub>1</sub> (21-99) was confirmed by ESI-FTICR MS/MS experiments (Fig. 2).**

Calculated monoisotop. [M+H] <sup>+</sup>	Uniprot Entry	Name (fragment)	Peptide sequences
6429.4127	P02654	ApoCI (29-83)	DVSSALDKLKEFGNTLEDKARELISRIKQSELSAKMREWFSETFQKVKEKLKIDS
6627.5131	P02654	<b>ApoCI (27-83)</b>	TPDVSSALDKLKEFGNTLEDKARELISRIKQSELSAKMREWFSETFQKVKEKLKIDS
7761.1843	P02776	Platelet Factor 4 (32-101)	EAEEDGDLQCLCVKTTSQVVRPRHITSLEVIKAGPHCPTAQLIATLKNRKCILDLQAPLYKKIKKLLLES
8200.0509	P02655	ApoCII (29-101)	DEMPSPTFLTQVKESLSSYWESAKTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMSTYTGIFTDQVLSVLKGEE
8689.1879	P02656	ApoCIII <sub>0a</sub> (21-98)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPTSAVA
8703.4457	P02652	ApoA2 (24-100)	QAKEPCVESLVSQYFQTVTDYGKDLMEKVKSPELQAEAKSYFEKSKEQLTPLIKKAGTELNVNLSYFVELGTQPATQ
8760.2250	P02656	ApoCIII <sub>0a</sub> (21-99)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPTSAVA
8910.3856	P02655	ApoCII (23-101)	TQQPQDEMPSTFLTQVKESLSSYWESAKTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMSTYTGIFTDQVLSVLKGEE
9054.3201	P02656	ApoCIII <sub>0c</sub> (21-98)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPT(GalNAcGal)SAVA
9125.3572	P02656	ApoCIII <sub>0c</sub> (21-99)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPT(GalNAcGal)SAVA
9282.8208	P02775	CTAP-III (44-128)	NLAGKKEESLSDSLYAEALRCMCIKTTSGIHPKNIQSLEVIKGGTHCNQVEVIATLKDGRKICLDPDAPRIKKIVQKLAGDESAD
9345.4155	P02656	ApoCIII <sub>1</sub> (21-98)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPT(GalNAcGalneu5Ac)SAVA
9416.4526	P02656	<b>ApoCIII<sub>1</sub> (21-99)</b>	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPT(GalNAcGalneu5Ac)SAVA
9707.5480	P02656	ApoCIII <sub>2</sub> (21-99)	SEAEDASLLSFMQGYMKHATKTKAKDALSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDDPEVRPT(GalNAcGalneu5AcNeu5Ac)SAVA
11676.5003	P02735	SAA1 (19-122)	RSFSSFLGEAFDGDARMWRAYSMDREANYIGSDKYFHARGNYDAAKRGPGGVWAAEAISDARENIQRFHGGAEDSLADQAANEWGRSGKDPNHFRA GLPEKY



**Figure 3.** Enlarged part of the ultrahigh resolution MALDI-FT-ICR protein profile obtained from human serum purified by RPC18 cartridge-based SPE. In this  $m/z$  range, Apo-CIII isoforms were identified as the most abundant peaks and Apo-CIII isoforms truncations (minus the C-terminal alanine), Apo-CII and the connective tissue-activating peptide III (CTAP-III) were observed at lower intensities. The upper inlets show the comparison between the observed isotopic distribution and those calculated from the amino acid sequences of the Apo-CIII isoforms (as indicated). The mass differences (365.13 and 291.10) between the Apo-CIII isoforms are in excellent agreement with the theoretical values (for HexNAcHex and NeuAc are 365.1322 and 291.0954 Da). The SDs of the mass differences were calculated from 96 RPC18 MB-based MALDI-FT-ICR protein profiles.

Furthermore, it should be stressed that the small glycoprotein Apo-CIII<sub>1</sub> can actually be a mixture of two isomers, namely the sialic acid residue can be linked to either the terminal Gal or to the GalNAc (NeuAc $\alpha$ 2-3Gal $\beta$ 1-3GalNAc or Gal $\beta$ 1-3[NeuAc $\alpha$ 2-6]GalNAc, respectively). The observed ESI-FT-ICR-MS/MS analysis is in full accordance with both structures (Fig. 2B). Therefore, no conclusions can be drawn from these data on the

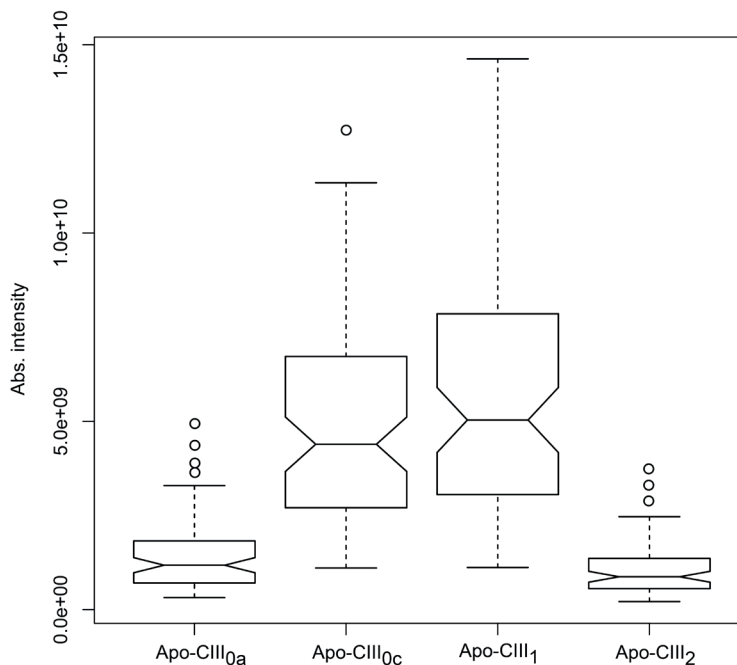
relative contribution of these two isomers to the Apo-CIII<sub>1</sub> species. Possibly electron transfer dissociation MS/MS experiments can differentiate between these two isomers [46]. As a final remark it becomes clear from MALDI-FT-ICR profiles that the mass difference between species nonglycosylated Apo-CIII, Apo-CIII<sub>0</sub>, Apo-CIII<sub>1</sub>, and Apo-CIII<sub>2</sub> provides additional evidence for the peak identities. This is illustrated in Fig. 3, where the observed  $\Delta m$  are 365.13, 291.10, and 291.10 Da, respectively (theoretical differences for HexNAcHex and NeuAc are 365.1322 and 291.0954 Da). Note that the mass difference of 0.02 Da corresponds to a MME of 2 ppm based on the mass of Apo-CIII<sub>1</sub> (calculated  $m/z$  9416.4526).

**Quantitative analysis of four apo-CIII isoforms.** In order to evaluate the biological variation in abundance of the apo-CIII isoforms in human sera, we analysed the MALDI-FT-ICR MS spectra obtained from 96 different individuals. A 96-well-plate containing RP18 MBs eluates and stored at  $-80^{\circ}\text{C}$  was thawed and each eluate was spotted in duplicate onto a MALDI target plate as described in Section 2. From this plate, 192 spectra were obtained and visualized using the DataAnalysis software (Bruker). First, the technical variation in the Apo-CIII isoform distributions was determined from each pair of spectra (replicate spots). Recently, we reported in more detail on the technical variation of similarly obtained MALDI-FT-ICR profiles [45, 47]. It was found that the reproducibility of peak areas of peptides and proteins between 1 and 9 kDa varied for each signal, with on average a coefficient of variation of approximately 15%. In the current profiles, similar technical variation was observed upon considering duplicate measurements of the same sample. The spectrum with highest intensities was used for further evaluation. In Fig. 3, a typical MALDI-FT-ICR serum protein profile in the mass range of 8550–9820  $m/z$  value is depicted. In this  $m/z$  range, Apo-CIII isoforms were identified as the most abundant peaks while Apo-CIII isoforms truncations (minus the C-terminal alanine), Apo-CII and the connective tissue-activating peptide III were identified as minor components (Table 1). The high confidence in the identifications derived from literature search or obtained from MS/MS analysis follows from the observed isotopic distribution of the Apo-CIII isoforms when compared with those calculated from the amino acid sequence using the IsotopePattern tool (Bruker). As shown in the upper insets of Fig. 3, the two observed and calculated isotopic distributions of all the Apo-CIII

isoforms have a good match with each other. Further note the excellent agreement of the mass differences between the Apo-CIII isoforms with the theoretical values. To calculate those mass differences, the  $m/z$  values of the most abundant isotope peak in the observed isotopic distribution of each of the Apo-CIII isoforms were extracted using the MassList tool (Bruker), copied in an Excel sheet and compared with the most abundant isotope peak in the calculated isotopic distribution of the corresponding Apo-CIII isoform. Thus, the average and SD of the mass differences in 96 spectra were calculated.

In order to evaluate the biological variation quantitative data on the abundances of the four species non-glycosylated Apo-CIII (Apo-CIII<sub>0a</sub>), Apo-CIII<sub>0c</sub>, Apo-CIII<sub>1</sub> and Apo-CIII<sub>2</sub> were obtained as follows. The signal intensity of the extracted  $m/z$  values was calculated using the Xtractor tool [16] and exported in an Excel sheet. These intensities are summarized in the box-plot depicted in Fig. 4. It is clear from this box-plot that significant variation is present in isoform abundances. This was confirmed through an ANOVA, which resulted in an F value of 87 with a p-value less than  $2 \times 10^{-16}$ . In view of the important role of apolipoproteins in various disease pathways and of the fact that the present analytical procedure allows high-throughput screening [45] while providing very detailed information, further studies are warranted. Therefore, future studies will be performed in which MALDI-FT-ICR profiling of the Apo-CIII isoforms will be applied in a clinical cohort. This requires a systematical evaluation of abundances and identities with respect to both biological and technical variations in a fully standardized setting.





**Figure 4. Box-plot of the signal intensities of Apo-CIII<sub>0a</sub>, Apo-CIII<sub>0c</sub>, Apo-CIII<sub>1</sub>, Apo-CIII<sub>2</sub>, as obtained from 96 ultrahigh resolution MALDI-FT-ICR protein profiles. Large biological variation in the abundance of the apolipoprotein-C isoforms in 96 different individuals is evident, ANOVA resulted in an F-value of 87 with a p-value less than  $2 \times 10^{-16}$ .**

## CONCLUSIONS

Peptide and protein profiles from human body fluids can be obtained from ultrahigh resolution MALDI instruments such as FT-ICR-MS with sub-ppm mass precision and four orders of magnitude in dynamic concentration range. Provided sample collection as well as SPE is performed according to highly standardized protocols these profiles hold great potential for screening (larger) cohorts of patient samples. Moreover, single species (*e.g.* proteins) in complex mixtures can be identified and quantified using ultrahigh resolution MS. In this study a fully automated RPC18-based magnetic bead method was applied for SPE of 96 different human serum samples. Proteins were profiled from  $m/z$  value 6000 to  $m/z$  value 15 000. The FT-ICR-system allowed accurate mass analysis

of middle-down proteins in MALDI protein profiles since all species were fully resolved in this mass range. It was shown that isoforms of apo-CIII could be mapped, namely the ones with the different mucin type core 1 O-glycans including one or two sialic acid residues. No evidence was found for the presence of Apo-CIII-GalNAc (“truncated” mucin-type core 1) in any of the acquired MALDI-FT-ICR profiles of all 96 different serum samples. Finally, ESI-FT-ICR-MS/MS identified the species observed in the MALDI-profile at  $m/z$  value 6627.5 as the monoisotopic peak of apo-CI.

## **ACKNOWLEDGEMENT**

The authors acknowledge funding from the research programme “Decrease Colorectal Cancer Death”, which is financially supported by the Center for Translational Molecular Medicine (CTMM). Furthermore they would like to thank Ekaterina Nevedomskaya for her support in statistical evaluations.

## REFERENCES

- [1] Albrethsen, J., The first decade of MALDI protein profiling: a lesson in translational biomarker research. *J. Proteomics* 2011, 74, 765–773.
- [2] Buchen, L., Missing the mark. *Nature* 2011, 471, 428–432.
- [3] Mitchell, P., Proteomics retrenches. *Nat. Biotech.* 2010, 28, 665–670.
- [4] Surinova, S., Schiess, R., Hüttenhain, R., Cerciello, F. et al., On the development of plasma protein biomarkers. *J. Proteome Res.* 2010, 10, 5–16.
- [5] Anderson, N. L., Counting the proteins in plasma. *Clin. Chem.* 2010, 56, 1775–1776.
- [6] Kulasingam, V., Pavlou, M. P., Diamandis, E. P., Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer* 2010, 10, 371–378.
- [7] de Noo, M. E., Tollenaar, R. A. E., Özalp, A., Kuppen, P. J. K. et al., Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal. Chem.* 2005, 77, 7232–7241.
- [8] Villanueva, J., Lawlor, K., Toledo-Crow, R., Tempst, P., Automated serum peptide profiling. *Nat. Protocols* 2006, 1, 880–891.
- [9] Albrethsen, J., Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clin. Chem.* 2007, 53, 852–858.
- [10] Callesen, A. K., Madsen, J. S., Vach, W., Kruse, T. A. et al., Serum protein profiling by solid phase extraction and mass spectrometry: a future diagnostics tool? *Proteomics* 2009, 9, 1428–1441.
- [11] Pan, S., Chen, R., Brand, R. E., Hawley, S. et al., Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. *J. Proteome Res.* 2012, 11, 1937–1948.
- [12] Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L. et al., A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* 2011, 29, 625–634.
- [13] Guan, S., Marshall, A. G., Scheppele, S. E., Resolution and chemical formula identification of aromatic hydrocarbons and aromatic compounds containing sulfur, nitrogen, or oxygen in petroleum distillates and refinery streams. *Anal. Chem.* 1996, 68, 46–71.
- [14] Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L. et al., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 2009, 8, 2759–2769.
- [15] Yates, J. R., Cociorva, D., Liao, L., Zabrouskov, V., Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* 2005, 78, 493–500.
- [16] Nicolardi, S., Palmblad, M., Hensbergen, P. J., Tollenaar, R. A. E. M. et al., Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* 2011, 25, 3457–3463.
- [17] Nedelkov, D., Phillips, D. A., Tubbs, K. A., Nelson, R. W., Investigation of human protein variants and their frequency in the general population. *Mol. Cell. Proteomics* 2007, 6, 1183–1187.
- [18] Tiss, A., Smith, C., Menon, U., Jacobs, I. et al., A well- characterised peak identification list of

- MALDI MS profile peaks for human blood serum. *Proteomics* 2010, 10, 3388–3392.
- [19] Hortin, G. L., The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.* 2006, 52, 1223–1237.
- [20] Jong, M. C., Hofker, M. H., Havekes, L. M., Role of ApoCs in lipoprotein metabolism: functional differences between ApoC1, ApoC2, and ApoC3. *Arterioscler. Thromb. Vasc. Biol.* 1999, 19, 472–484.
- [21] Saland, J., Ginsberg, H. N., Lipoprotein metabolism in chronic renal insufficiency. *Ped. Nephrol.* 2007, 22, 1095–1112.
- [22] Gordon, S. M., Deng, J., Lu, L. J., Davidson, W. S., Proteomic characterization of human plasma high density lipoprotein fractionated by gel filtration chromatography. *J. Proteome Res.* 2010, 9, 5239–5249.
- [23] Shachter, N. S., Apolipoproteins C-I and C-III as important modulators of lipoprotein metabolism. *Curr. Opin. Lipidol.* 2001, 12, 297–304.
- [24] Sacks, F. M., Zheng, C., Cohn, J. S., Complexities of plasma apolipoprotein C-III metabolism. *J. Lipid Res.* 2011, 52, 1067–1070.
- [25] Mauger, J.-F., Couture, P., Bergeron, N., Lamarche, B., Apolipoprotein C-III isoforms: kinetics and relative implication in lipid metabolism. *J. Lipid Res.* 2006, 47, 1212–1218.
- [26] Wang, C. S., McConathy, W. J., Kloer, H. U., Alaupovic, P., Modulation of lipoprotein-lipase activity by apolipoproteins — effect of apolipoprotein-C-III. *J. Clin. Invest.* 1985, 75, 384–390.
- [27] Davidsson, P., Hulthe, J., Fagerberg, B., Olsson, B.-M. et al., A proteomic study of the apolipoproteins in LDL subclasses in patients with the metabolic syndrome and type 2 diabetes. *J. Lipid Res.* 2005, 46, 1999–2006.
- [28] Batal, R., Tremblay, M., Barrett, P. H. R., Jacques, H. et al., Plasma kinetics of apoC-III and apoE in normolipidemic and hypertriglyceridemic subjects. *J. Lipid Res.* 2000, 41, 706–718.
- [29] Onat, A., Hergenc, G., Sansoy, V., Fobker, M. et al., Apolipoprotein C-III, a strong discriminant of coronary risk in men and a determinant of the metabolic syndrome in both genders. *Atherosclerosis* 2003, 168, 81–89.
- [30] Jensen, M. K., Rimm, E. B., Furtado, J. D., Sacks, F. M., Apolipoprotein C-III as a potential modulator of the association between HDL-cholesterol and incident coronary heart disease. *J. Am. Heart Assoc.* 2012, 1, 1–10.
- [31] Cohen, M., Yossef, R., Erez, T., Kugel, A. et al., Serum apolipoproteins C-I and C-III are reduced in stomach cancer patients: results from MALDI-based peptidome and immuno-based clinical assays. *PLoS One* 2011, 6, e14540.
- [32] Wopereis, S., Grünewald, S., Morava, E., Penzien, J. M. et al., Apolipoprotein C-III isofocusing in the diagnosis of genetic defects in O-glycan biosynthesis. *Clin. Chem.* 2003, 49, 1839–1845.
- [33] Wopereis, S., Grünewald, S., Huijben, K. M. L. C., Morava, E. et al., Transferrin and apolipoprotein C-III isofocusing are complementary in the diagnosis of N and O-glycan biosynthesis defects. *Clin. Chem.* 2006, 53, 180–187.
- [34] Brousseau, T., Clavey, V., Bard, J.-M., Fruchart, J.-C., Sequential ultracentrifugation micromethod for separation of serum-lipoproteins and assays of lipids, apolipoproteins, and lipoprotein particles. *Clin. Chem.* 1993, 39, 960–964.

- [35] Sprecher, D. L., Taam, L., Brewer, H. B., Two-dimensional electrophoresis of human-plasma apolipoproteins. *Clin. Chem.* 1984, 30, 2084–2092.
- [36] Visvikis, S., Dumon, M.-F., Steinmetz, J., Manabe, T. et al., Plasma apolipoproteins in tangier disease, as studied with two-dimensional electrophoresis. *Clin. Chem.* 1987, 33, 120–122.
- [37] Bruneel, A., Morelle, W., Carre, Y., Habarou, F. et al., Twodimensional gel electrophoresis of apolipoprotein C-III and other serum glycoproteins for the combined screening of human congenital disorders of a and N-glycosylation. *Proteomics Clin. Appl.* 2007, 1, 321–324.
- [38] Haase, R., Menke-Möllers, I., Oette, K., Analysis of human apolipoproteins-C by isoelectric-focusing in immobilized pH gradients. *Electrophoresis* 1988, 9, 569–575.
- [39] Hancock, W. S., Bishop, C. A., Gotto, A. M., Harding, D. R. K et al., Separation of the apoprotein components of human very low-density lipoproteins by ion-paired, reversed-phase high-performance liquid-chromatography. *Lipids* 1981, 16, 250–259.
- [40] Mazur, M. T., Cardasis, H. L., Spellman, D. S., Liaw, A. et al., Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc. Nat. Acad. Sci. USA* 2010, 107, 7728–7733.
- [41] Nelsestuen, G. L., Zhang, Y., Martinez, M. B., Key, N. S. et al., Plasma protein profiling: unique and stable features of individuals. *Proteomics* 2005, 5, 4012–4024.
- [42] Jimenez, C. R., El Filali, Z., Knol, J. C., Hoekman, K. et al., Automated-serum peptide profiling using novel magnetic C18 beads off-line coupled to MALDI-TOF-MS. *Proteomics Clin. Appl.* 2007, 1, 598–604.
- [43] Wada, Y., Kadoya, M., Okamoto, N., Mass spectrometry of apolipoprotein C-III, a simple analytical method for mucintype O-glycosylation, and its application to an autosomal recessive cutis laxa type-2 (ARCL2) patient. *Glycobiology* 2012, 22, 1140–1144.
- [44] Nicolardi, S., Palmblad, M., Dalebout, H., Bladergroen, M. et al., Quality control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spectrom.* 2010, 21, 1515–1525.
- [45] Bladergroen, M. R., Derks, R. J. E., Nicolardi, S., de Visser, B. et al., Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids. *J. Proteomics* 2012, 77, 144–153.
- [46] Nicolardi, S., Dalebout, H., Bladergroen, M. R., Mesker, W. E. et al., Identification of human serum peptides in Fourier transform ion cyclotron resonance precision profiles. *Int. J. Proteomics* 2012, 6, 804036.
- [47] Selman, M. H. J., McDonnell, L. A., Palmblad, M., Ruhaak, L. R., Deelder, immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 2010, 82, 1073–1081.





# Chapter 6

Identification of New Apolipoprotein-CIII  
Glycoforms with Ultrahigh Resolution  
MALDI-FTICR Mass Spectrometry  
of Human Sera

*Simone Nicolardi, Yuri E. M. van der Burgt,  
Irina Dragan, Paul J. Hensbergen,  
and André M. Deelder*

J. Proteome Res.  
2013, 12, 2260–2268



## **ABSTRACT**

Apolipoprotein-CIII (apoCIII) is an abundant blood glycoprotein associated with lipoprotein particles. Three different glycoforms have been described, all containing a mucin-type core-1 O-glycosylation with either zero, one or two sialic acids. Changes in the relative abundance of these glycoforms have been observed in a variety of different pathologies. In this study, ultrahigh resolution 15T MALDI Fourier transform ion cyclotron resonance (FTICR) MS was used to analyze apoCIII isoforms in serum protein profiles. For this purpose, serum proteins were purified using both a fully automated RPC18-based magnetic bead method and an RPC4 cartridge-based solid phase extraction method. Six new apoCIII isoforms were identified with low-ppm mass measurement errors and ultrahigh precision. These were characterized by more complex glycan moieties which are fucosylated instead of sialylated. To confirm the glycan moiety and localize the glycosylation site top-down ESI-FTICR-MS/MS and bottom-up LC-ion trap MS/MS were used. A large variation in the presence and abundance of the fucosylated isoforms was found in a set of 96 serum samples. These findings of fucosylated apoCIII isoforms warrant further research to elucidate the implications these glycoforms may have for the plethora of studies where alterations in apoCIII have been linked to the development of many different pathologies.

## INTRODUCTION

Apolipoprotein-CIII (apoCIII) is a relatively small protein containing 79 amino acids. It is present in blood where it is associated with chylomicrons, very-low-density lipoproteins (VLDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs) [1]. ApoCIII regulates the metabolism of lipoprotein particles through the inhibition of lipoprotein lipase thereby affecting the concentration of triacylglycerol in blood [2,3]. Therefore, high levels of apoCIII correlate with hypertriglyceridemia [4] and are associated with coronary heart disease and atherosclerosis [5]. ApoCIII has also direct pro-inflammatory effects on both monocytes and endothelial cells which may additionally contribute to atherogenic mechanisms [6]. Recently, a role for apoCIII in the development of type I diabetes has been demonstrated [7] and correlations between quantitative changes in the apoCIII levels and other diseases like ischemic and hemorrhagic stroke and stomach cancer have been described [8,9].

ApoCIII is glycosylated at threonine-74 and three glycoforms have been described [10]. All share a mucin-type core-1 *O*-glycosylation, *i.e.* a galactose  $\beta$ 1-3 linked to an *N*-acetylgalactosamine (Gal $\beta$ 1-3GalNAc). Two isoforms are further modified with either one or two sialic acid residues, respectively (Gal $\beta$ 1-3[NeuAc $\alpha$ 2-6]GalNAc and NeuAc $\alpha$ 2-3Gal $\beta$ 1-3[NeuAc $\alpha$ 2-6]GalNAc) [11]. The different apoCIII glycoforms are commonly referred to as apoCIII<sub>0c</sub>, apoCIII<sub>1</sub> and apoCIII<sub>2</sub>, the index indicating the number of sialic acid residues. A small fraction of apoCIII is non-glycosylated (apoCIII<sub>0a</sub>) [10]. An abnormal profile of the different apoCIII glycoforms (*i.e.* hypoglycosylation) can be used for diagnosis of genetic defects in *O*-glycan biosynthesis [12,13]. Because of the association of apoCIII levels with many pathological conditions, there is considerable interest to measure the levels of apoCIII and its glycoforms in serum. For this purpose, several methods have been developed including isoelectric focusing, two-dimensional electrophoresis, high-performance liquid chromatography and matrix-assisted laser desorption ionization (MALDI) coupled with time-of-flight (TOF) mass spectrometry (MS) [14-21].

In a few MS-based profiling studies novel serum apoCIII glycoforms have been suggested but the exact nature of the glycan composition remained undetermined [22,23]. Further support for the presence of additional apoCIII glycoforms comes from a recent study performed in our laboratory, in which aberrantly *O*-glycosylated apoCIII-derived peptides

were observed in urine from *Schistosoma mansoni*-infected individuals that were absent in non-infected individuals [24]. Although these peptides were presumably also *O*-glycosylated on Thr-74 (referring to the full length protein), the glycans on these peptides were completely different from the hitherto described *O*-glycosylation. Most importantly, these *O*-glycopeptides contained high levels of fucosylation but lacked sialylation.

Recently, we developed a new high-throughput MALDI Fourier transform ion cyclotron resonance (FTICR) MS method for profiling of serum proteins with fully resolved isotopic distributions up to 15300 Da [25]. Novel aspects of this method are the improved resolving power, precision and mass accuracy of the mass measurements, resulting in a more robust and reliable comparison of multiple spectra. The application of this method resulted in a detailed map of the previously reported apoCIII glycoforms in human serum [25].

In the current study, this MALDI-FTICR protein profiling method was used to explore the presence of additional apoCIII glycoforms in human serum. Both top-down and bottom-up experiments were performed on partially fractionated apoCIII by direct infusion ESI-FTICR MS/MS and liquid chromatography (LC) ion trap (IT) MS/MS, respectively. This resulted in the identification of six new apoCIII glycoforms in human serum. These are characterized by more complex glycan moieties which are fucosylated instead of sialylated. Furthermore, the relative distribution of apoCIII glycoforms was evaluated in a set of 96 different samples.

## **MATERIALS AND METHODS**

**Sample collection.** Human blood samples were obtained as previously described [26]. Briefly, blood samples were collected in 10 mL BD Vacutainer tubes (containing a clot activator and a gel for serum separation) from healthy volunteers by antecubital venipuncture and centrifuged within 1 h. Informed consent was obtained from all volunteers and the study was approved by the Medical Ethical Committee of the LUMC. Serum was then transferred to a 1 mL cryovial and stored at -80°C. For aliquoting, 96 cryovials were thawed and serum was dispensed in a 96-tube rack using an eight-channel liquid-handling robot (Hamilton, Bonaduz, Switzerland). This rack was then stored again at -80°C until the automated peptide isolation procedure was performed.

**High-throughput serum protein RP-C18 solid-phase extraction.** Serum samples were anonymized before protein purification by reversed-phase extraction (SPE) based on magnetic beads (MB's) [27]. SPE was carried out as follows: the RPC18-beads were first activated by a three-step washing with a 0.1% trifluoroacetic acid (TFA) solution. Then, 5  $\mu\text{L}$  of each sample were added to the activated beads and incubated for 5 min at room temperature. The beads were washed again three times with 0.1% TFA and proteins were eluted with a 1:1 mixture of water and acetonitrile. The MALDI spotting was performed on the liquid handling robot by mixing 2  $\mu\text{L}$  of the sample eluates with 10  $\mu\text{L}$  of an  $\alpha$ -cyano-4-hydroxycinnamic acid solution (0.3 g/L in ethanol/acetone 2:1). Each sample was spotted in duplicate onto a MALDI AnchorChip (600  $\mu\text{m}$ ; Bruker Daltonics, Bremen, Germany).

**Serum protein RP-C4 solid-phase extraction.** For identification purposes, (apolipo)proteins were first fractionated from serum samples by acetonitrile (ACN) precipitation followed by RP-C4 SPE. Briefly, 100  $\mu\text{L}$  of thawed human serum were diluted with 100  $\mu\text{L}$  of deionized water in a 1.5 mL plastic tube (Eppendorf) and kept at room temperature for 15 min for standardization reasons. High molecular weight proteins were precipitated by adding 200  $\mu\text{L}$  of ice-cold acetonitrile (ACN) and storing the tube for two hours at  $-20^{\circ}\text{C}$ . After 15 min at room temperature, the sample was centrifuged (Eppendorf, Centrifuge 5415 R) at 16,100 rcf for 10 min. Then, 300  $\mu\text{L}$  of the supernatant were transferred to a new 1.5 mL Eppendorf tube and diluted with 1200  $\mu\text{L}$  of a 0.05% formic acid (FA) solution. For each sample four replicate extractions were performed and the total volume (6 mL) of depleted-diluted serum was loaded on a 1 ml BioSelect SPE C4 cartridge (Grace, USA) that was pre-washed twice with 900  $\mu\text{L}$  of a 95% ACN / 4.95% water / 0.05% FA solution and pre-conditioned with 900  $\mu\text{L}$  of a 0.05% FA solution. The cartridge was then washed three times with 900  $\mu\text{L}$  of a 10% ACN / 89.95% water / 0.05% FA solution and the adsorbed peptides and proteins were stepwise eluted with 900  $\mu\text{L}$  of a 15%, 20%, 25% and 30% acetonitrile solution. Prior to ESI-FTICR MS, 60  $\mu\text{L}$  of each eluate was diluted with 40  $\mu\text{L}$  of a 95% acetonitrile / 4.95% water / 0.05% FA solution and analyzed by direct infusion ESI-FTICR MS.

**MALDI- and ESI-FTICR mass spectrometry.** Both MALDI- and ESI-FTICR experiments were performed on a Bruker 15 Tesla solariX™ FTICR mass spectrometer equipped with a CombiSource as previously reported [25,28]. Briefly, the MALDI-FTICR system was equipped with a Bruker Smartbeam-II™ Laser System that operated at a frequency of 1000 Hz. Each mass spectrum was acquired from  $m/z$ -values 6,000 to 15,300 and obtained from the average of 10 scans of 150 laser shots each using 256 K data points. The random walk option was allowed on a diameter of 300  $\mu\text{m}$ . The ions were trapped and measured in the ICR cell using the Gated Trapping option. Both the front and the back trapping potentials were set to 1 V while the front and the back detection voltages were set to 0.4 and 0.45 V, respectively. The ramp time was 0.01 s. The required excitation power was 45% with a pulse time of 15.0  $\mu\text{s}$ .

Prior to the measurement of each MALDI plate the FTICR system was externally calibrated using a commercially available peptide mix and a protein mix (Bruker Daltonics). The spectra were then internally calibrated using the  $m/z$ -values 6631.5242, 8765.2387, 8915.3993, 9130.3710, 9421.4644, 11729.7978 corresponding to the most intense isotope peak of apolipoprotein-CI (amino acids 27-83), apoCIII<sub>0a</sub> (21-99), apoCIII<sub>0a</sub> (23-99), apoCIII<sub>0c</sub> (21-99), apoCIII<sub>1</sub> (21-99), and beta-2-microglobulin (21-19), respectively. DataAnalysis Software 4.0 SP 3 (Bruker Daltonics) was used for the visualization and the calibration of the spectra. After evaluation by manual inspection, the spectrum with highest signal intensities was selected from each duplicate spectra for further statistical analysis. Both the  $m/z$ -values and intensity values were visualized using DataAnalysis Software 4.0 SP 3 (Bruker Daltonics) and copied in a Microsoft Excel worksheet for further evaluation. Theoretical isotopic distributions were calculated using the Simulate Isotopic Pattern tool using similar resolving powers as obtained in the protein profiles. Direct infusion ESI-FTICR experiments were performed using the same instrumental settings described previously with some modification [28]. Briefly, a quadrupole (Q) was used for precursor ion selection and a hexapole collision cell for collision-induced dissociation (CID). Direct infusion electrospray ionisation (ESI) experiments were carried out at an infusion rate of 2  $\mu\text{L}/\text{min}$ . The ion funnels operated at 100 V and 6.0 V, respectively, with the skimmers at 15 V and 5 V. The trapping potentials were set at 0.60 V and 0.55 V, the analyzer entrance was maintained at -7 V, and side kick technology was used to further optimize peak shape and signal intensity.

The required excitation power was 19% with a pulse time of 10  $\mu$ s. MS/MS-experiments were performed by CID and fragment ion mass analysis in the ICR cell. For these experiments, the collision energy, the accumulation time in the hexapole collision cell and the isolation window in the Q were optimized for each precursor ion. Collision energies varied from 5 to 23 V while the accumulation times varied from 1 to 10 s.

**Trypsin Digestion and LC-Ion Trap Mass Spectrometry.** Serum samples were processed as described above and 300  $\mu$ L of the 20% ACN eluate from the RP-C4-cartridge was lyophilized. Proteins were subsequently reconstituted in 50  $\mu$ L ammonium bicarbonate pH 7.9 and digested with trypsin. For this purpose, proteins were first reduced and alkylated using dithiothreitol (5 mM, 30 min, 56°C) followed by iodoacetamide (15 mM, 30 min, RT, in the dark). Then, 50 ng trypsin (sequencing grade modified Trypsin, Promega, Madison, WI) was added and proteins were digested overnight at 37°C. The digestion reaction was stopped using 2  $\mu$ L of 5% TFA. The digests were then analyzed by LC-MS using an Ultimate 3000 RSLCnano LC system (Thermo Scientific, Sunnyvale, CA) coupled to an HCTultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The samples were injected onto an Acclaim C18 PepMap100 trapping column (100  $\mu$ m $\times$ 2 cm, 5  $\mu$ m, 100 Å (Thermo Scientific) and washed with 100% A (3% ACN in 0.1% formic acid) at 5  $\mu$ L/min for 6 min. Following valve switching, peptides were separated on an Acclaim C18 PepMapRSLC column (75  $\mu$ m $\times$ 150 mm, 2  $\mu$ m, 100 Å) at a constant flow of 300 nL/min. The peptide elution gradient was from 3 to 40 % B (95% ACN in 0.1% formic acid) in 48 min followed by an increase to 65% B in 10 min. The nanoflow LC was coupled to the mass spectrometer using a nano-electrospray ionization source. The spray voltage was set at 1.2 kV and the temperature of the heated capillary was set to 165°C. Eluting peptides were fragmented by CID using helium as the collision gas.

## RESULTS

**MALDI-FTICR-MS profiling of serum apoCIII glycoforms.** Recently, we developed a method for fully automated RPC18 magnetic beads-based 15T MALDI-FTICR-MS protein profiling in the mass range from 6000 to 15300 Da [25]. These spectra were found to be dominated by various apolipoproteins. In the current study, we used this method to explore the presence of new apoCIII glycoforms in serum samples of 96 healthy control individuals. Firstly, all hitherto characterized apoCIII (glyco)forms (apoCIII<sub>0a</sub>, apoCIII<sub>0c</sub>, apoCIII<sub>1</sub> and apoCIII<sub>2</sub>) could be mapped in all 96 different samples with our ultrahigh resolution MALDI-FTICR protein profiling method (Table 1). Upon focusing in-depth on ions with low intensities in the mass range from 9690 to 10900 Da (Figure 1) it was noted that these spectra contained  $m/z$ -values that corresponded to the theoretical ones for the monoprotonated apoCIII isoforms assuming the same *O*-glycan structures as were identified on apoCIII-derived peptides in urines of *Schistosoma mansoni*-infected individuals, as overviewed in Table 1 [24]. These glycoforms have a more complex and fucosylated glycan structure and do not contain any sialic acid residues.

It should be stressed that, although the monoisotopic peaks are generally not observed in these profiles, the isotopic distribution itself of each apoCIII isoform allows the accurate determination of its mass (Figure 1). Thus, a nominal mass difference of 1 Da, *e.g.* the difference between two Fuc residues and one sialic acid, would be observable (see Figure S1 in the Supplementary Materials). This is further illustrated in Figure S1, S2 and S3 in the Supplementary Materials, where theoretical distributions are compared to observed ones and the most abundant isotope peak within a distribution is used for the calculation of the mass measurement error. The observed low mass measurement errors (MMEs) corroborated the identification of each glycoform [28 29]. The MMEs were in the low ppm range (0.2 – 3 ppm) in all 96 different samples; however it should be noted that in none of the profiles (corresponding to a single sample) *all* six different glycoforms were observed at the same time. If one would show only the “best examples”, the observed MME of each glycoform would be at a sub-ppm level. In all cases, *i.e.* for all different species in all 96 different serum profiles, the mass differences between the “known” apoCIII isoforms and “newly identified” ones was determined with very high precision ( $\pm 0.020$  Da).

**Table 1. Isoforms of apolipoprotein-CIII**

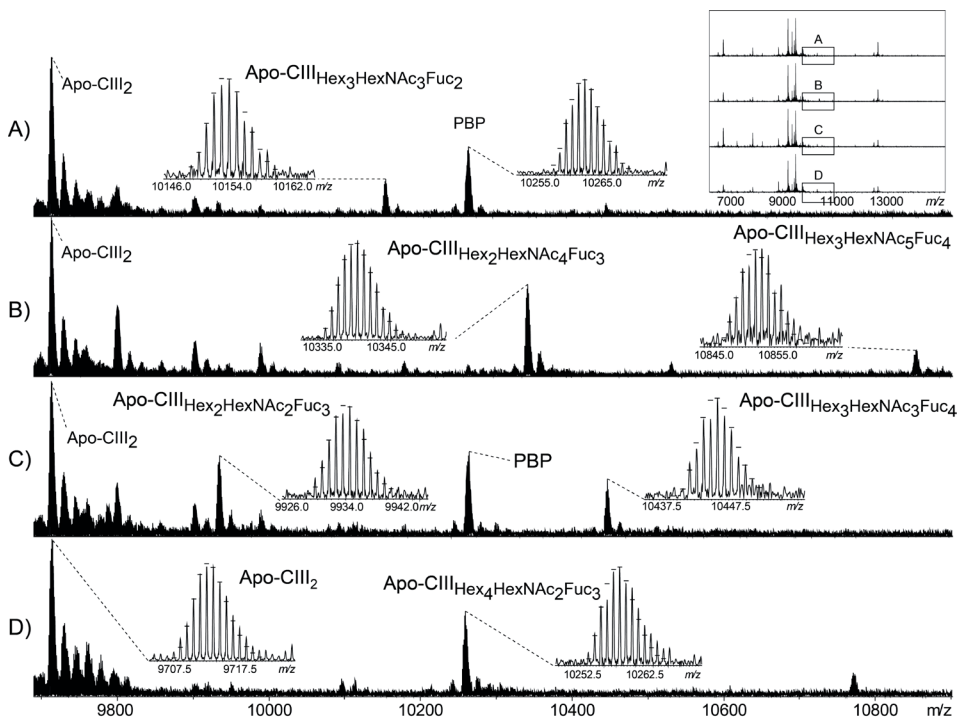
Apolipoprotein-CIII amino acid sequence: SEAEDASLLSFMQGYMKHATKTAKDA LSSVQESQVAQQARGWVTDGFSSLKDYWSTVKDKFSEFWLDLPEVRPTSAVAA		
Observed apolipoprotein-CIII isoform	Calculated m/z-value [M+H] <sup>+</sup>	Frequency in 96 serum samples
apoCIII <sub>0a</sub> (-)	8765.2385	100%
apoCIII <sub>0c</sub> (HexHexNAc)	9130.3708	100%
apoCIII <sub>1</sub> (HexHexNAcNeuAc) <sup>a</sup>	9421.4663	100%
apoCIII <sub>2</sub> (HexHexNAcNeuAc <sub>2</sub> ) <sup>b</sup>	9712.5618	100%
apoCIII-Hex <sub>2</sub> HexNAc <sub>2</sub> Fuc <sub>3</sub> <sup>b, c</sup>	9933.6769	27%
apoCIII-Hex <sub>3</sub> HexNAc <sub>3</sub> Fuc <sub>2</sub>	10153.7538	11%
apoCIII-Hex <sub>4</sub> HexNAc <sub>2</sub> Fuc <sub>3</sub>	10258.7852	8%
apoCIII-Hex <sub>2</sub> HexNAc <sub>4</sub> Fuc <sub>3</sub>	10340.8383	29%
apoCIII-Hex <sub>3</sub> HexNAc <sub>3</sub> Fuc <sub>4</sub>	10445.8698	12%
apoCIII-Hex <sub>3</sub> HexNAc <sub>5</sub> Fuc <sub>4</sub>	10852.0286	6%

<sup>a</sup> Structure confirmed with ESI-FTICR-MS/MS in a previous study [25].

<sup>b</sup> Structure confirmed with ESI-FTICR-MS/MS.

<sup>c</sup> Structure confirmed with LC-IT-MS/MS.





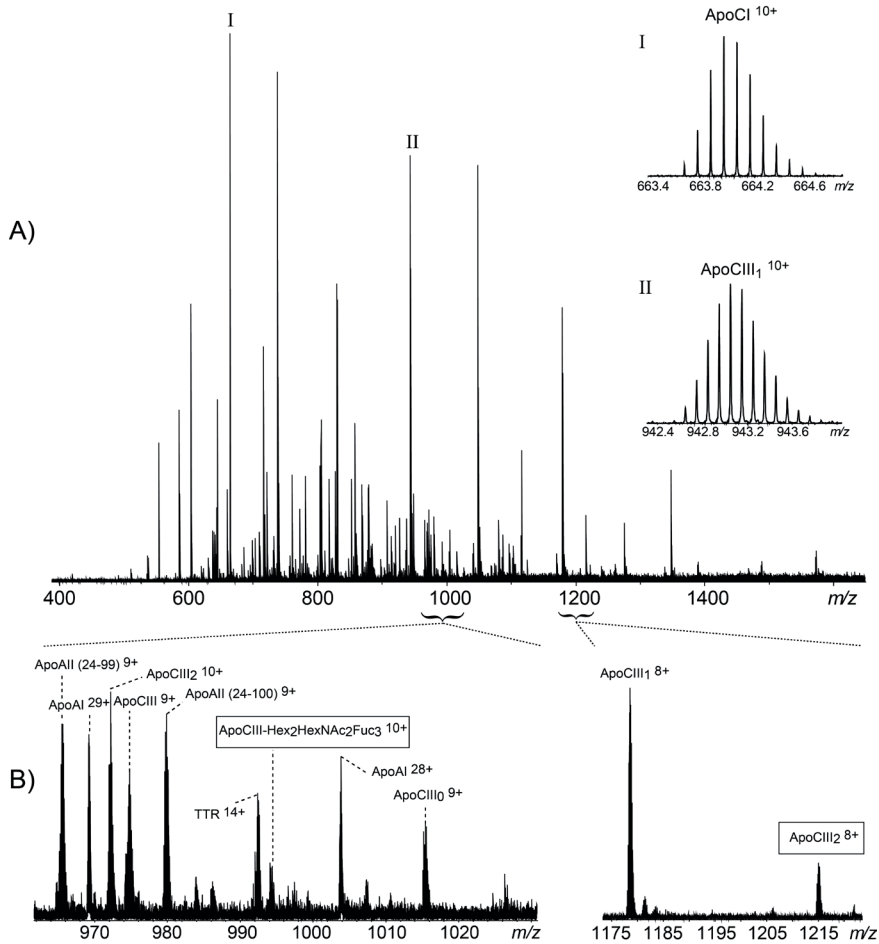
**Figure 1.** Top-down 15T MALDI-FTICR-MS identification of fucosylated apolipoprotein-CIII isoforms in human serum. Human serum proteins were purified by RP-C18 magnetic bead-based SPE and profiled by MALDI-FTICR MS from  $m/z$ -value 6000 up to  $m/z$ -value 15300 (top-right inset). In the enlarged part of four different protein profiles (A–D), isotopically resolved apolipoprotein-CIII glycoforms are shown in the  $m/z$ -range from 9690 to 10900. Six fucosylated apolipoprotein-CIII isoforms were identified by accurate mass measurement and were observed with different frequency in the spectra obtained from different individuals (see Table 1). The observed isotopic distributions are in good agreement with the isotopic distribution calculated from the theoretical molecular formulas. Hex, Hexose; HexNAc, N-acetylhexosamine; Fuc, Fucose; PBP, platelet basic protein.

In all 96 serum profiles, a large variation in the relative abundance of all the apoCIII glycoforms was observed. However, there was not only variation in the relative abundance of these species, but also in the frequency wherein a specific isoform was observed. (Table1). For example, the most frequently observed fucosylated apoCIII isoform (apoCIII-Hex<sub>2</sub>HexNAc<sub>4</sub>Fuc<sub>3</sub>) was found in 29% of the samples while apoCIII-Hex<sub>3</sub>HexNAc<sub>5</sub>Fuc<sub>4</sub> was present in only 6% of the samples. For relative quantification purposes, we normalized the signal intensity of the fucosylated isoforms to that of apoCIII<sub>0a</sub>. As such, the relative signal intensities ranged from 9% (isoform apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>) to 3% (isoform apoCIII-Hex<sub>3</sub>HexNAc<sub>5</sub>Fuc<sub>4</sub>) with relative standard deviations (RSD) of 3% and 1%, respectively.

In summary, we have shown that the combination of ultrahigh mass measurement precision and resolution, with concomitant low MMEs, allows the identification of new proteoforms [29]. As such, six multi-fucosylated apoCIII isoforms were identified in human serum samples. One of these contains a glycan with three hexoses, three *N*-acetylhexosamines and four fucoses (*i.e.* apoCIII-Hex<sub>3</sub>HexNAc<sub>3</sub>Fuc<sub>4</sub>) that was not reported in the previous study on urinary apoCIII-derived peptides from *Schistosoma mansoni*-infected individuals.

**Top-Down Identification of apoCIII glycoforms by direct infusion ESI-FTICR mass spectrometry.** To further confirm the identification of intact forms of apoCIII, top-down MS/MS experiments were performed on fractionated apolipoprotein samples. For this purpose, serum samples were first depleted from high molecular weight proteins using organic precipitation. These samples were loaded on C4-cartridges, and the eluates obtained from the stepwise elution of serum peptides/proteins with 15, 20, 25, 30% acetonitrile solutions were measured by direct infusion ESI-FTICR-MS. The spectra obtained from the first three fractions (15, 20 and 25% acetonitrile) showed similar protein charge state distributions while the spectrum of the fourth fraction was different and dominated by the charge state distribution of apolipoprotein AI (ApoAI, data not shown). The apoCIII<sub>0a</sub>, apoCIII<sub>0c</sub>, apoCIII<sub>1</sub> and apoCIII<sub>2</sub> isoforms were observed with maximum signal intensity in the spectrum obtained from the second fraction. This spectrum is depicted in Figure 2a. Here, the most intense peaks were identified as apolipoprotein CI (apoCI) and apoCIII<sub>1</sub>. The exact *m/z*-value of the most abundant isotope

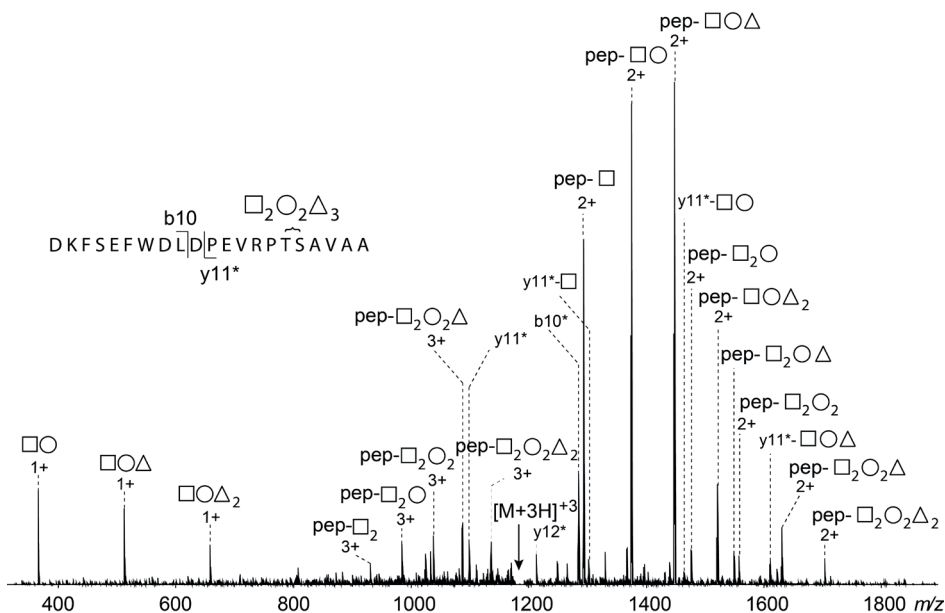
peak in the isotopic distributions of these two proteins was used to internally calibrate the ESI-FTICR MS spectrum. The enlarged parts of the ESI-FTICR spectrum (Figure 2b) show ion species ( $m/z$  1215.1 ( $[M+8H]^{8+}$ ) and  $m/z$  994.3 ( $[M+10H]^{10+}$ )) corresponding to apoCIII<sub>2</sub> and to apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>. Subsequently, CID experiments were performed on the ion at  $m/z$  1215.1 ( $[M+8H]^{8+}$ ), of which the results are depicted in Figure 3a. Similar to the MS spectra, the isotopic distribution of each theoretical ion was used to assign CID fragments in MS<sup>2</sup>-spectra. Clear losses of two sialic acids, one hexose and one *N*-acetylhexosamine are observed. Moreover, the singly charged ions observed in the mass range from  $m/z$ -values 250 to 1095 match with the oxonium ions corresponding to the glycan composition of the apoCIII<sub>2</sub> isoform (*i.e.* Hex-HexNAc-Neu5Ac<sub>2</sub>). In addition to the glycosidic bond cleavages, protein backbone fragments, such as the b<sub>77-</sub> and b<sub>78-</sub>-ions with or without one *N*-acetylhexosamine, were observed. The CID spectrum of the ion at  $m/z$  994.3 ( $[M+10H]^{10+}$ ), corresponding to apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>, is depicted in Figure 3b. Here, losses of three fucoses, two hexoses and two *N*-acetylhexosamines unambiguously identify the glycan moiety on this protein species. This is further supported by the oxonium ions observed in the mass range from  $m/z$  450 to 850. In addition, the backbone fragmentation is similar to that of the apoCIII<sub>2</sub> isoform described above although in this case the b<sub>77-</sub> and the b<sub>78-</sub>-ions are only observed with an *N*-acetylhexosamine. Finally, two fragment ions containing one hexose, one *N*-acetylhexosamine and three fucoses were observed. These ions most likely result from fucose rearrangements that can occur during CID fragmentation of multi-fucosylated glycopeptides [30].



**Figure 2.** Ultrahigh resolution 15T ESI-FTICR-MS of human serum apolipoproteins. (A) ESI-FTICR MS spectrum of human serum apolipoproteins fractionated by RP-C4 cartridge-based SPE after the removal of large proteins by acetonitrile precipitation. The two most abundant proteins in the spectrum were identified as apoCI and apoCIII<sub>1</sub>. (B) Enlarged parts of the ESI-FTICR spectrum show less abundant proteins including ion species (*i.e.*,  $m/z$  1215.1,  $[M + 8]^{8+}$ ;  $m/z$  994.3,  $[M + 10]^{10+}$ ) corresponding to apoCIII<sub>2</sub> and to apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>. These two species were further analyzed by top-down 15T ESI-FTICR MS/MS (see Figure 3). Hex, Hexose; HexNAc, N-acetylhexosamine; Fuc, Fucose; TTR, Transthyretin.



In the top-right inset, the singly charged oxonium ions corresponding to the singly protonated glycan moieties are shown. All assigned CID fragments are reported in Table S1 in Supporting Information. (B) CID-spectrum of apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub> ( $m/z$ -value 994.3;  $[M + 10]^{10+}$ ) obtained with direct infusion ESI-FTICR-MS/MS of serum samples after acetonitrile precipitation and RPC4 cartridge-based SPE (see Figure 2). Most of the intense peaks were identified as fragments resulting from the losses of fucoses, hexoses and *N*-acetylhexosamines from the glycan moiety (*i.e.*, Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>). Two fragment ions were identified as b<sub>77</sub><sup>\*</sup> and b<sub>78</sub><sup>\*</sup> ions with one *N*-acetylhexosamine, respectively. Three singly charged oxonium ions corresponding to singly protonated glycans were observed (top-right inset). All assigned CID fragments are reported in Table S2 in Supporting Information. Square and HexNAc, *N*-acetylhexosamine; circle, hexose; triangle; fucose; diamond, *N*-acetylneuraminic acid; b<sub>77</sub><sup>\*</sup>, b<sub>77</sub> without glycans; b<sub>78</sub><sup>\*</sup>, b<sub>78</sub> without glycans.



**Figure 4.** Bottom-Up LC-IT-MS identification of a fucosylated apolipoprotein-CIII isoform. CID-spectrum of a tryptic peptide from apoCIIIHex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub> ( $m/z$ -value 1183.5;  $[M + 3]^{3+}$ ) obtained after liquid chromatography and ion trap MS/MS. Multiply charged ions were identified as fragments resulting from the loss of fucoses, hexoses and *N*-acetylhexosamines. Three singly charged oxonium ions corresponding to singly protonated glycans were observed in the low-mass region. One N-terminal fragment ion (*i.e.*, b<sub>10</sub>) was observed without the

glycosylation while one C-terminal fragment ion (*i.e.*,  $y_{11}^*$  with either a HexNAc, HexNAc-Hex or HexNAc-Hex-Fuc moiety) was observed. All assigned CID fragments are reported in Table S3 in Supporting Information. Square and HexNAc, N-acetylhexosamine; circle, hexose; triangle; fucose.  $y_{11}^*$ ,  $y_{11}$  without glycans.

**Bottom-up identification by LC-IT Mass Spectrometry.** To further substantiate the identification of fucosylated apoCIII isoforms in serum as derived from our MALDI-FTICR MS and top-down ESI-FTICR MS analyses, a bottom-up approach was chosen. To this end, tryptic peptides were obtained from the RP-C4 20% acetonitrile fraction (see above) and these were analyzed by LC-IT mass spectrometry. An inclusion list was used to limit the CID experiments to precursor ions corresponding to one of the glycosylated tryptic apoCIII peptides (DKFSEFWLDLPEVRPTSAVAA-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>), related to one of the fucosylated apoCIII isoforms identified by our top-down analysis (Figure 3b). The CID spectrum obtained from the fragmentation of the triply-charged precursor ion at  $m/z$  1183.5 is depicted in Figure 4. Clear losses of the fucoses, hexoses and N-acetylhexosamines are apparent. In addition, oxonium ions corresponding to the HexNAc-Hex element containing one or two fucoses were observed. Some peptide backbone cleavages were also apparent. In line with what has been observed previously [24] the most prominent fragments correspond to the cleavage between the aspartic acid and proline residue. The resulting C-terminal fragment ions without ( $y_{11}^*$  at  $m/z$ -value 1097.6) and with either a HexNAc, HexNAc-Hex or HexNAc-Hex-Fuc moiety were clearly noticeable. This demonstrates that, similar to the known apoCIII glycoforms, the glycosylation resides in the C-terminal part of apoCIII, most probably also on Thr-74.

## DISCUSSION

Apolipoprotein-CIII is one of the abundant serum glycoproteins and has been biochemically characterized approximately 40 years ago [2,10,11,31]. In these initial studies it already became evident that there are several different apoCIII glycoforms which are characterized by a different degree of sialylation. At that time, the apoCIII glycoforms were separated and analyzed using laborious and time consuming techniques which were later substituted by more elaborate and fast analytical methods, including mass spectrometry [1,15-17]. In the current study, we have used a combination of

ultrahigh resolution MALDI-FTICR MS, top-down ESI-FTICR-MS/MS and a bottom-up approach on tryptic glycopeptides to study the *O*-glycosylation of apoCIII in serum. In addition to the known glycoforms, this resulted in the characterization of new low abundant apoCIII isoforms, containing more complex, fucosylated glycan moieties, possibly similar to Lewis-type  $\text{Fuca}(1-2)\text{Gal}\beta(1-4)[\text{Fuca}(1-3)]\text{GlcNAc}$  or  $\text{Gal}\beta(1-4)[\text{Fuca}(1-3)]\text{GlcNAc}$  structures.

Previously, MALDI-TOF and SELDI-TOF based methods have been used to study plasma/serum apolipoprotein profiles [8,9]. Although these techniques are very valuable to assign known peptides and proteins, these MS-systems lack the necessary resolution, mass accuracy and precision for a confident characterization of new species as is exemplified by earlier reports on apoCIII isoforms [22,23]. Here, the putatively identified novel apoCIII species at 9934 Da probably corresponds to the most prominent and frequently observed fucosylated apoCIII isoform (apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub>) that we characterized in the current study. The application of our ultrahigh resolution MALDI-FTICR method allowed the identification of six fucosylated apoCIII isoforms based on low-ppm precision of multiple profiles and corresponding low MMEs [29,32]. With increasing number of applications on an orbitrap system this approach is also coined as “high resolution/accurate mass (HR/AM)”. A top-down ESI-FTICR-MS/MS was used for further characterization of the new apoCIII isoforms. The CID-analysis on the apoCIII-Hex<sub>2</sub>HexNAc<sub>2</sub>Fuc<sub>3</sub> isoform primarily resulted in glycosidic bond cleavages, which confirmed the glycan moiety. However, relatively few peptide backbone fragments were observed. This is a common phenomenon of CID-mediated glycopeptide/protein fragmentation and often hampers the correct assignment of the glycosylation site [33]. Other fragmentation techniques like electron-transfer dissociation (ETD) can overcome this problem provided the intensity of the precursor ion is sufficient for performing such experiments (which unfortunately was not the case for the currently reported species) [17,34]. We therefore have applied a bottom-up approach to localize the site of glycosylation. Although based on our data the exact position (Thr-74 or Ser-75) could not be inferred, we strongly believe that, in line with previous reports, Thr-74 is the site of glycosylation.

There has been significant interest for profiling different apoCIII glycoforms due to the association of apoCIII with many (patho)physiological conditions, for example



triglyceride levels [3] and the risk of cardiovascular diseases [5,6]. In the future, it would be indicated to include the newly identified fucosylated species in the analyses of ultrahigh resolution protein profiles and determine the relation between their quantitative changes and, for example, cardiovascular physiology. For the current study, all samples were controls from an in-house study cohort and there appeared to be a large variation in the presence of the fucosylated apoCIII isoforms, both qualitatively as well as quantitatively. Considering the quantification, no firm conclusion about the relative quantities of the different isoforms can be drawn from our data, primarily because we have no information about the relative ionization efficiency of the different isoforms. In spite of this, it appeared that the fucosylated isoforms are not the dominating isoforms. Notwithstanding, signal intensities of 50% of the major fucosylated isoform relative to the apoCIII<sub>2</sub> have been frequently observed (Figure 1). Most of the glycan structures on the fucosylated serum apoCIII isoforms were previously found on apoCIII-derived peptides in urine of *Schistosoma mansoni*-infected individuals but not in urine from healthy individuals [24]. In that study, the full-length fucosylated isoforms were not detected using MALDI-TOF analysis, neither in the serum from infected nor in the serum from healthy individuals. Our new method opens possibilities for further exploration, in order to reveal whether the appearance of the peptides in the urine of infected individuals is related to altered serum levels of these glycoforms. Alternatively, it may be the result of specific proteolytic activity associated with *Schistosoma mansoni*-infection in combination with the incomplete removal of these degradation products in the urinary tract system.

As stated above, the novel glycan structures of apoCIII most probably are attached to the same residue (Thr-74) as the hitherto identified apoCIII glycosylation. Mutagenesis of this site revealed that apoCIII glycosylation is not indispensable for the secretion and distribution of apoCIII over the different lipoprotein particles [35] and individuals carrying a Thr-74 to Ala-74 mutation are apparently asymptomatic although they showed elevated levels of apoCIII<sub>0</sub> [36]. ApoCIII inhibits lipoprotein lipase (LPL) and the importance of *O*-glycosylation in relation to this was recently demonstrated in a study of a loss-of-function mutant in the GalNAc transferase *GALNT2* [37]. ApoCIII is specifically glycosylated by this enzyme and the mutation results in higher serum levels of apoCIII<sub>0</sub> and lower levels of apoCIII<sub>1</sub> and is associated with accelerated clearance of plasma

triglycerides. In accordance, desialylation of apoCIII abrogates the LPL-inhibitory effect. At present, we cannot make any prediction on the functional role of the newly identified apoCIII glycoforms but we envisage that their identification is essential for obtaining a comprehensive assessment on the structure-function relationship of this major serum glycoprotein.

In conclusion, using a combination of MS-techniques, we have characterized a set of fucosylated apoCIII isoforms in serum. In comparison to the known mucin-type, core-1 apoCIII glycosylation with different degrees of sialylation, these novel glycoforms are characterized by more complex, fucosylated glycan moieties. The here presented analytical protein profiling procedure allows high-throughput screening while providing very detailed information. This in combination with the fact that apolipoproteins play an important role in various disease pathways warrants further research. Here, we aim to elucidate the implications these glycoforms may have for the plethora of studies where alterations in apoCIII have been linked to the development of many different pathologies.

## SUPPORTING INFORMATION

**Table 1. Fragment ions observed in the CID ESI-FTICR spectrum of ApoCIII<sub>2</sub>**

Fragment Ion	Observed most abundant isotope peak; <i>m/z</i> (charge state)	Calculated most abundant isotope peak; <i>m/z</i> (charge state)	$\Delta$ Mass (Da)	Mass Measurement Error (ppm)
ApoCIII <sub>2</sub> *	1214.9519 (8+)	292.1027 (1+)	0.0001	0.34
SA	292.1028 (1+)	495.1821 (1+)	-0.0001	-0.20
SAHexNAc	495.1820 (1+)	657.2349 (1+)	-0.0001	-0.15
SAHexNAcHex	657.2348 (1+)	948.3303 (1+)	0.0000	0.00
SA <sub>2</sub> HexNAcHex	948.3303 (1+)	1096.5362 (8+)	0.0023	2.1
ApoCIII0a	1096.5385 (8+)	1253.0403 (7+)	0.0001	0.08
	1253.0404 (7+)	1461.7125 (6+)	-0.0003	-0.21
	1461.7122 (6+)	1110.7901 (8+)	0.0009	0.81
b78(HexNAc)	1110.7910 (8+)	1101.9105 (8+)	0.0000	0.00
b77(HexNAc)	1101.9105 (8+)	1121.9211 (8+)	0.0007	0.62
ApoCIII(HexNAc)	1121.9218 (8+)	1282.0517 (7+)	0.0001	0.08
	1282.0518 (7+)	1495.5591 (6+)	-0.0005	-0.33
	1495.5586 (6+)	1142.1777 (8+)	0.0005	0.44
ApoCIII(HexNAcHex)	1142.1782 (8+)	1305.2021 (7+)	0.0002	0.15
	1305.2023 (7+)	1522.5679 (6+)	0.0001	0.07
	1522.5680 (6+)	1178.5647 (8+)	0.0002	0.17
ApoCIII(HexNAcHexSA)	1178.5649 (8+)	1346.7871 (7+)	-0.0004	-0.30
	1346.7867 (7+)	1214.9516 (8+)	0.0003	0.25
b78	1240.3197 (7+)	1240.3192 (7+)	0.0005	0.40
	1446.8709 (6+)	1446.8712 (6+)	-0.0003	-0.21
b77	1230.1713 (7+)	1230.1711 (7+)	0.0002	0.16
	1435.0317 (6+)	1435.0317 (6+)	0.0000	0.00
ApoCIII(HexNAcSA)	1158.3086 (8+)	1158.308 (8+)	0.0006	0.52
	1323.6351 (7+)	1323.6367 (7+)	-0.0016	-1.2

\* Precursor Ion

**Table 2. Fragment ions observed in the CID ESI-FTICR spectrum of apoCIII(HexNAc<sub>2</sub>Hex<sub>2</sub>Fuc<sub>3</sub>)**

Fragment Ion	Observed most abundant isotope peak;	Calculated most abundant isotope peak;	$\Delta$ Mass (Da)	Mass Measurement Error (ppm)
apoCIII(HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc <sub>3</sub> )*	994.2743 (10+)	994.2742 (10+)	-0.0001	-0.10
HexHexNAcFuc	512.1974 (1+)	512.1974 (1+)	0.0000	0.00
HexHexNAcFuc <sub>2</sub>	658.2553 (1+)	658.2553 (1+)	0.0000	0.00
HexHexNAcFuc <sub>3</sub>	804.3129 (1+)	804.3132 (1+)	0.0003	0.37
apoCIII	974.8105 (9+)	974.8108 (9+)	0.0003	0.31
	1096.5365 (8+)	1096.5362 (8+)	-0.0003	-0.27
apoCIII(2HexNAc)	918.0461 (10+)	918.0463 (10+)	0.0002	0.22
apoCIII(HexNAc <sub>2</sub> Hex)	934.2515 (10+)	934.2516 (10+)	0.0001	0.11
apoCIII(HexNAc <sub>2</sub> Hex <sub>2</sub> )	950.4569 (10+)	950.4569 (10+)	0.0000	0.00
apoCIII(HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc)	965.0630 (10+)	965.0627 (10+)	-0.0003	-0.31
apoCIII(HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc <sub>2</sub> )	979.6678 (10+)	979.6684 (10+)	0.0006	0.61
apoCIII(HexNAc)	997.3752 (9+)	997.3751 (9+)	-0.0001	-0.10
	1121.9214 (8+)	1121.9211 (8+)	-0.0003	-0.27
apoCIII(HexNAcHex)	1015.3812 (9+)	1015.381 (9+)	-0.0002	-0.20
	1142.1777 (8+)	1142.1777 (8+)	0.0000	0.00
apoCIII(HexNAcHexFuc)	1031.6095 (9+)	1031.6097 (9+)	0.0002	0.19
	1160.4355 (8+)	1160.435 (8+)	-0.0005	-0.43
apoCIII(HexNAcHexFuc <sub>2</sub> )	1047.8380 (9+)	1047.8383 (9+)	0.0003	0.29
apoCIII(HexNAcHexFuc <sub>3</sub> )	1064.0669 (9+)	1064.067 (9+)	0.0001	0.09
b78(HexNAc)	1110.7908 (8+)	1110.7901 (8+)	-0.0007	-0.63
b77(HexNAc)	1101.9080 (8+)	1101.9105 (8+)	0.0025	2.27

\* Precursor Ion

**Figure 3. Fragment ions observed in the CID Ion Trap MS/MS spectrum of a tryptic peptide of apoCIII(HexNAc<sub>2</sub>Hex<sub>2</sub>Fuc<sub>3</sub>)**

Fragment Ion	Observed most abundant isotope peak; <i>m/z</i> (charge state)	Calculated most abundant isotope peak; <i>m/z</i> (charge state)	$\Delta$ Mass (Da)	Mass Measurement Error (ppm)
Pep*-HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc <sub>3</sub> **	1183.5 (3+)**	1183.53 (3+)**	0.03	25
HexHexNAc	366.10 (1+)	366.14 (1+)	0.04	109
HexHexNAcFuc	512.16 (1+)	512.20 (1+)	0.04	78
HexHexNAcFuc <sub>2</sub>	658.24 (1+)	658.26 (1+)	0.02	30
Pep-HexNAc	1292.61 (2+)	1292.62 (2+)	0.01	8
Pep-HexNAc <sub>2</sub>	929.81 (3+)	929.77 (3+)	-0.04	-43
Pep-HexNAcHex	1373.61 (2+)	1373.64 (2+)	0.03	22
Pep-HexNAcHexFuc	1446.65 (2+)	1446.67 (2+)	0.02	14
Pep-HexNAcHexFuc <sub>2</sub>	1519.60 (2+)	1519.70 (2+)	0.10	66
Pep-HexNAc <sub>2</sub> Hex	983.76 (3+)	983.79 (3+)	0.03	30
	1475.08 (2+)	1475.18 (2+)	0.10	68
Pep-HexNAc <sub>2</sub> HexFuc	1548.08 (2+)	1548.21 (2+)	0.13	84
Pep-HexNAc <sub>2</sub> Hex <sub>2</sub>	1037.77 (3+)	1037.81 (3+)	0.04	39
	1556.14 (2+)	1556.21 (2+)	0.07	45
Pep-HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc	1086.48 (3+)	1086.5 (3+)	0.02	18
	1629.12 (2+)	1629.24 (2+)	0.12	74
Pep-HexNAc <sub>2</sub> Hex <sub>2</sub> Fuc <sub>2</sub>	1135.14 (3+)	1135.18 (3+)	0.04	35
	1702.13 (2+)	1702.27 (2+)	0.14	82
b10	1283.58 (1+)	1283.56 (1+)	-0.02	-16
y11*	1097.61 (1+)	1097.59 (1+)	-0.02	-18
y12*	1212.55 (1+)	1212.62 (1+)	0.07	58
y11*HexNAc	1300.63 (1+)	1300.67 (1+)	0.04	31
y11*HexNAcHex	1462.65 (1+)	1462.73 (1+)	0.08	55
y11*HexNAcHexFuc	1608.65 (1+)	1608.79 (1+)	0.14	87

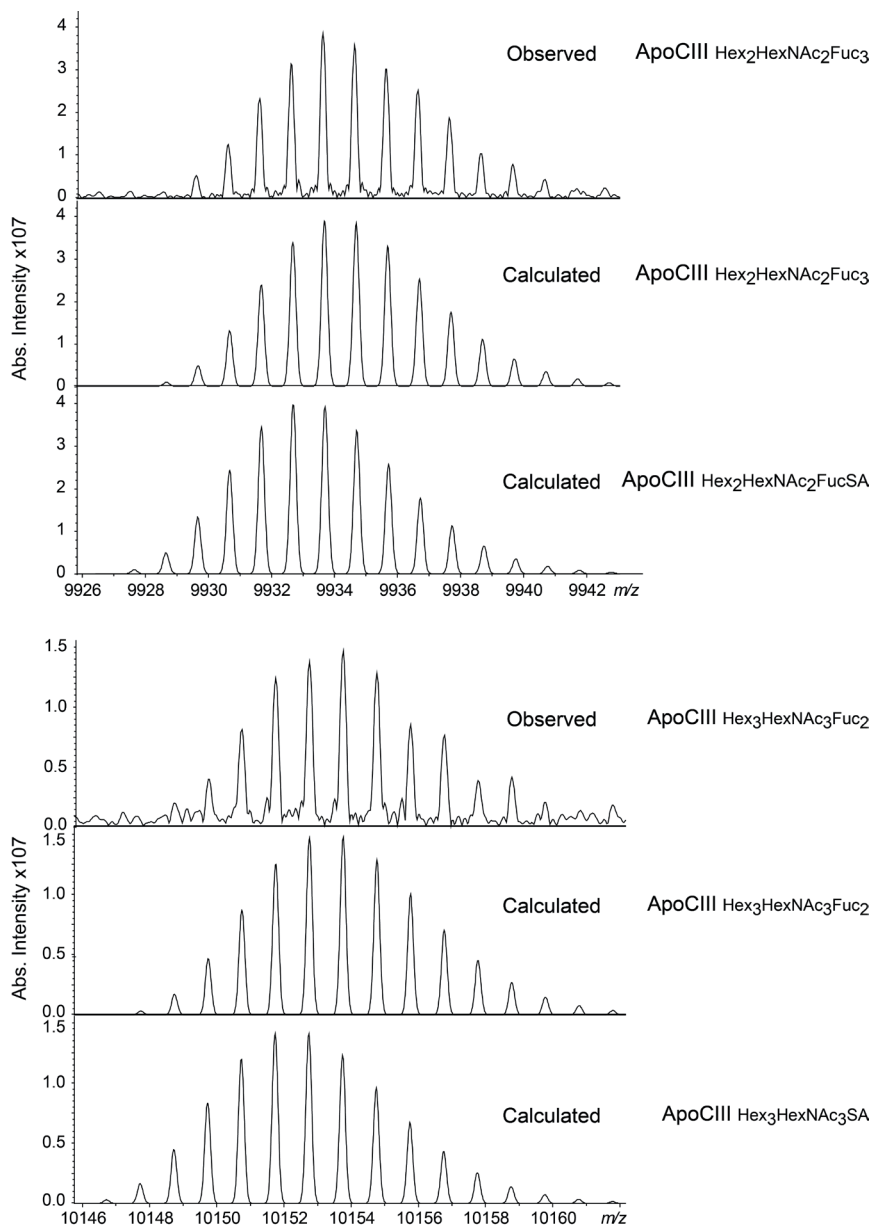
\* Pep: DKFSEFWDLDPV RPTSAVAA

\*\* Precursor Ion

\*\*\* Monoisotopic peak

**Figure S1, S2 and S3.** Enlarged parts of MALDI-FTICR serum protein profiles. The observed and theoretical isotopic distribution is shown for each fucosylated apoCIII isoform and for beta2microglobulin.

**Figure S1**



**Figure S2**

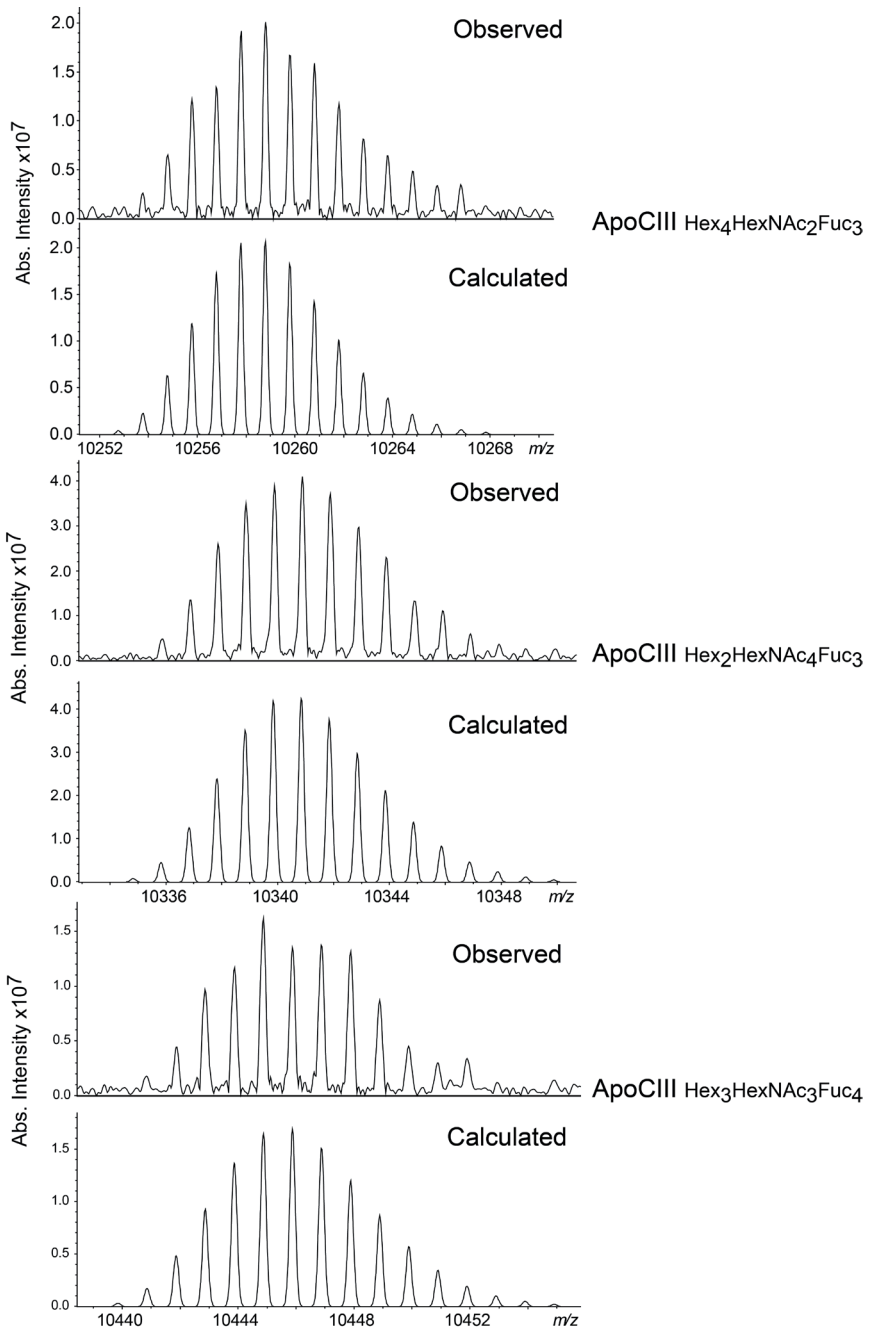
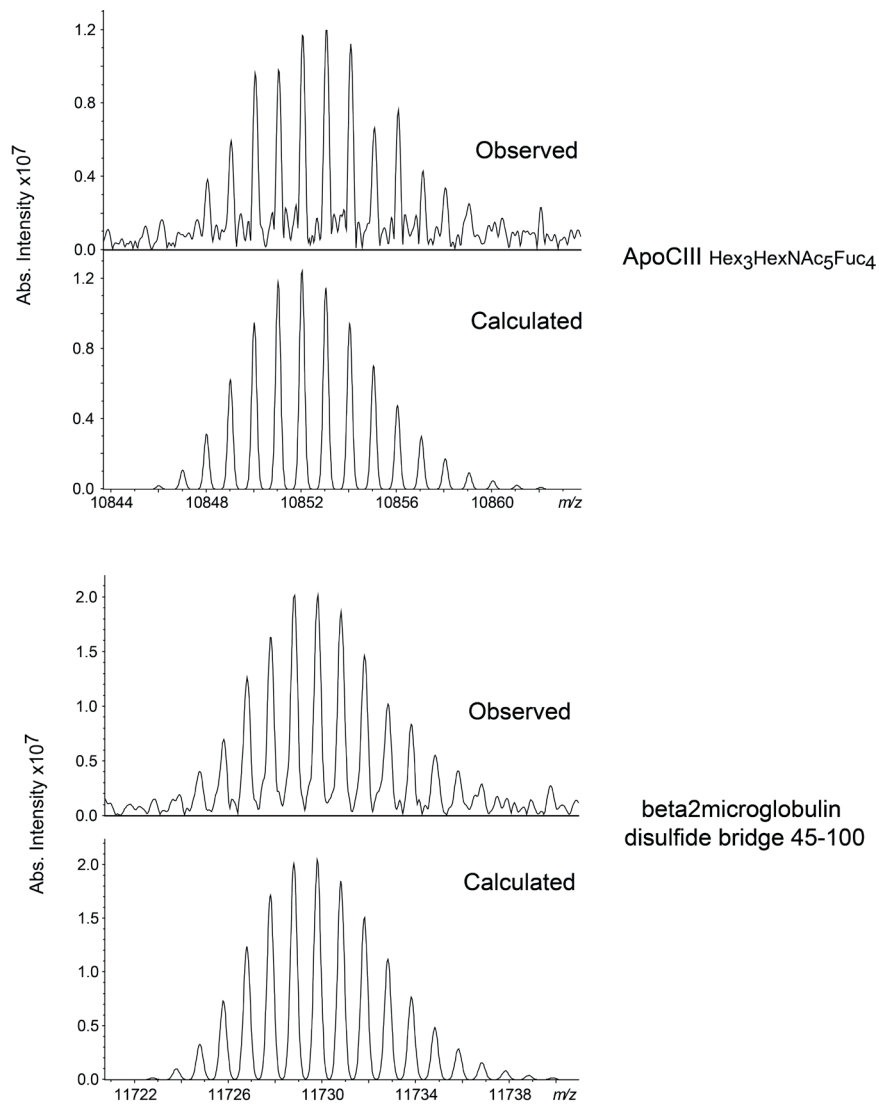


Figure S3





## REFERENCES

1. Jong, M. C.; Hofker, M. H.; Havekes, L. M. Role of ApoCs in lipoprotein metabolism - Functional differences between ApoC1, ApoC2, and ApoC3. *Arterioscler. Thromb. Vasc. Biol.* 1999, 19 (3), 472-484.
2. Mauger, J.-F.; Couture, P.; Bergeron, N.; Lamarche, B. Apolipoprotein C-III isoforms: kinetics and relative implication in lipid metabolism. *J. Lipid Res.* 2006, 47 (6), 1212-1218.
3. Wang, C. S.; McConathy, W. J.; Kloer, H. U.; Alaupovic, P. Modulation of lipoprotein-lipase activity by apolipoproteins - Effect of apolipoprotein-C-III. *J. Clin. Invest.* 1985, 75 (2), 384-390.
4. Batal, R.; Tremblay, M.; Barrett, P. H. R.; Jacques, H.; Fredenrich, A.; Mamer, O.; Davignon, J.; Cohn, J. S. Plasma kinetics of apoC-III and apoE in normolipidemic and hypertriglyceridemic subjects. *J. Lipid Res.* 2000, 41 (5), 706-718.
5. Onat, A.; Hergenc, G.; Sansoy, V.; Fobker, M.; Ceyhan, K.; Toprak, S.; Assmann, G. Apolipoprotein C-III, a strong discriminant of coronary risk in men and a determinant of the metabolic syndrome in both genders. *Atherosclerosis* 2003, 168 (1), 81-89.
6. Kawakami, A.; Yoshida, M. Apolipoprotein CIII Links Dyslipidemia with Atherosclerosis. *J. Atheroscler. Thromb.* 2009, 16 (1), 6-11.
7. Juntti-Berggren, L.; Refai, E.; Appelskog, I.; Andersson, M.; Imreh, G.; Dekki, N.; Uhles, S.; Yu, L.; Griffiths, W. J.; Zaitsev, S.; Leibiger, I.; Yang, S. N.; Olivecrona, G.; Jönrnvall, H.; Berggren, P. O. Apolipoprotein CIII promotes Ca<sup>2+</sup>-dependent  $\beta$  cell death in type 1 diabetes. *Proc. Nat. Acad. Sci. USA* 2004, 101 (27), 10090-10094.
8. Allard, L.; Lescuyer, P.; Burgess, J.; Leung, K.-Y.; Ward, M.; Walter, N.; Burkhard, P. R.; Corthals, G.; Hochstrasser, D. F.; Sanchez, J.-C. ApoC-I and ApoC-III as potential plasmatic markers to distinguish between ischemic and hemorrhagic stroke. *Proteomics* 2004, 4 (8), 2242-2251.
9. Cohen, M.; Yossef, R.; Erez, T.; Kugel, A.; Welt, M.; Karpasas, M. M.; Bones, J.; Rudd, P. M.; Taieb, J.; Boissin, H.; Harats, D.; Noy, K.; Tekoah, Y.; Lichtenstein, R. G.; Rubin, E.; Porgador, A. Serum apolipoproteins C-I and C-III are reduced in stomach cancer patients: results from MALDI-based peptidome and immuno-based clinical assays. *PLoS ONE* 2011, 6 (1), e14540.
10. Vaith, P.; Assmann, G.; Uhlenbruck, G. Characterization of the oligosaccharide side chain of apolipoprotein C-III from human plasma very low density lipoproteins. *BBA-Gen Subjects* 1978, 541 (2), 234-240.
11. Brewer, H. B.; Shulman, R.; Herbert, P.; Ronan, R.; Wehrly, K. The Complete Amino Acid Sequence of Alanine Apolipoprotein (apoC-III), an Apolipoprotein from Human Plasma Very Low Density Lipoproteins. *J. Biol. Chem.* 1974, 249 (15), 4975-4984.
12. Wopereis, S.; Grünewald, S.; Morava, É.; Penzien, J. M.; Briones, P.; García-Silva, M. T.; Demacker, P. N. M.; Huijben, K. M. L. C.; Wevers, R. A. Apolipoprotein C-III isofocusing in the diagnosis of genetic defects in O-glycan biosynthesis. *Clin. Chem.* 2003, 49 (11), 1839-1845.
13. Wopereis, S.; Grünewald, S.; Huijben, K. M. L. C.; Morava, É.; Mollicone, R.; van

- Engelen, B. G. M.; Lefeber, D. J.; Wevers, R. A. Transferrin and apolipoprotein C-III isofocusing are complementary in the diagnosis of N- and O-glycan biosynthesis defects. *Clin. Chem.* 2006, 53 (2), 180-187.
14. Sprecher, D. L.; Taam, L.; Brewer, H. B. Two-dimensional electrophoresis of human-plasma apolipoproteins. *Clin. Chem.* 1984, 30 (12), 2084-2092.
15. Bruneel, A.; Morelle, W.; Carre, Y.; Habarou, F.; Dupont, D.; Hesbert, A.; Durand, G.; Michalski, J. C.; Drouin-Garraud, V.; Seta, N. Two-dimensional gel electrophoresis of apolipoprotein C-III and other serum glycoproteins for the combined screening of human congenital disorders of a and N-glycosylation. *Proteomics Clin. Appl.* 2007, 1 (3), 321-324.
16. Hancock, W. S.; Bishop, C. A.; Gotto, A. M.; Harding, D. R. K.; Lamplugh, S. M.; Sparrow, J. T. Separation of the apoprotein components of human very low-density lipoproteins by ion-paired, reversed-phase high-performance liquid-chromatography. *Lipids* 1981, 16 (4), 250-259.
17. Mazur, M. T.; Cardasis, H. L.; Spellman, D. S.; Liaw, A.; Yates, N. A.; Hendrickson, R. C. Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc. Nat. Acad. Sci.* 2010, 107(17), 7728-7733.
18. Wada, Y.; Kadoya, M.; Okamoto, N. Mass spectrometry of apolipoprotein C-III, a simple analytical method for mucin-type O-glycosylation, and its application to an autosomal recessive cutis laxa type-2 (ARCL2) patient. *Glycobiology* 2012, 22 (8), 1140-1144.
19. Zhang, Y.; Sinaiko, A.; Nelsestuen, G. Glycoproteins and Glycosylation: Apolipoprotein C3 Glycoforms by Top-Down MALDI-TOF Mass Spectrometry. In *Liver Proteomics in Methods in Molecular Biology*, 909 ed.; Humana Press: 2012; pp 141-150.
20. Haase, R.; Menke-Möllers, I.; Oette, K. Analysis of human apolipoproteins-C by isoelectric-focusing in immobilized pH gradients. *Electrophoresis* 1988, 9 (9), 569-575.
21. Madian, A. G.; Regnier, F. E. Profiling Carbonylated Proteins in Human Plasma. *J. Proteome Res.* 2010, 9 (3), 1330-1343.
22. Bondarenko, P. V.; Cockrill, S. L.; Watkins, L. K.; Cruzado, I. D.; Macfarlane, R. D. Mass spectral study of polymorphism of the apolipoproteins of very low density lipoprotein. *J. Lipid Res.* 1999, 40 (3), 543-555.
23. Nelsestuen, G. L.; Zhang, Y.; Martinez, M. B.; Key, N. S.; Jilma, B.; Verneris, M.; Sinaiko, A.; Kasthuri, R. S. Plasma protein profiling: Unique and stable features of individuals. *Proteomics* 2005, 5 (15), 4012-4024.
24. Balog, C. I. A.; Mayboroda, O. A.; Wuhrer, M.; Hokke, C. H.; Deelder, A. M.; Hensbergen, P. J. Mass Spectrometric Identification of Aberrantly Glycosylated Human Apolipoprotein C-III Peptides in Urine from *Schistosoma mansoni*-infected Individuals. *Mol. Cell. Proteomics* 2010, 9 (4), 667-681.
25. Nicolardi, S.; Yuri E.M.van der Burgt; Wuhrer, M.; Deelder, A. M. Mapping O-glycosylation of apolipoprotein C-III in MALDI-FTICR protein profiles. *Proteomics* 2012, doi: 10.1002/pmic.201200293
26. Nicolardi, S.; Palmblad, M.; Dalebout, H.; Bladergroen, M.; Tollenaar, R. A. E. M.; Deelder, A. M.; van der Burgt, Y. E. M. Quality

control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spectrom.* 2010, 21 (9), 1515-1525.

27. Bladergroen, M. R.; Derks, R. J. E.; Nicolardi, S.; de Visser, B.; van Berloo, S.; van der Burgt, Y. E. M.; Deelder, A. M. Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids. *J. Proteomics* 2012, 77, 144-153.

28. Nicolardi, S.; Dalebout, H.; Bladergroen, M. R.; Mesker, W. E.; Tollenaar, R. A. E. M.; Deelder, A. M.; van der Burgt, Y. E. M. Identification of Human Serum Peptides in Fourier Transform Ion Cyclotron Resonance Precision Profiles. Identification of Human Serum Peptides in Fourier Transform Ion Cyclotron Resonance Precision Profiles. *Int. J. Proteomics* 2012, 804036

29. He, F.; Emmett, M. R.; Håkansson, K.; Hendrickson, C. L.; Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *J. Proteome Res.* 2003, 3 (1), 61-67.

30. Wührer, M.; Deelder, A. M.; van der Burgt, Y. E. M. Mass spectrometric glycan rearrangements. *Mass Spectrom. Rev.* 2011, 30 (4), 664-680.

31. Morrisett, J. D.; Jackson, R. L.; Gotto, A. M. Lipoproteins: Structure and Function. *Annu. Rev. Biochem.* 1975, 44 (1), 183-207.

32. Wu, S.; Tao, N.; German, J. B.; Grimm, R.; Lebrilla, C. B. Development of an

Annotated Library of Neutral Human Milk Oligosaccharides. *J. Proteome Res.* 2010, 9 (8), 4138-4151.

33. Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B* 2007, 849, 115-128.

34. Thaysen-Andersen, M.; Wilkinson, B. L.; Payne, R. J.; Packer, N. H. Site-specific characterisation of densely O-glycosylated mucin-type peptides using electron transfer dissociation ESI-MS/MS. *Electrophoresis* 2011, 32 (24), 3536-3545.

35. Roghani, A.; Zannis, V. I. Mutagenesis of the glycosylation site of human ApoCIII. O-linked glycosylation is not required for ApoCIII secretion and lipid binding. *J. Biol. Chem.* 1988, 263 (34), 17925-17932.

36. Maeda, H.; Hashimoto, R. K.; Ogura, T.; Hiraga, S.; Uzawa, H. Molecular cloning of a human apoC-III variant: Thr 74→Ala 74 mutation prevents O-glycosylation. *J. Lipid Res.* 1987, 28 (12), 1405-1409.

37. Holleboom, A.; Karlsson, H.; Lin, R. S.; Beres, T.; Sierts, J.; Herman, D.; Stroes, E.; Aerts, J.; Kastelein, J.; Motazacker, M.; Dallinga-Thie, G.; Levels, J.; Zwinderman, A.; Seidman, J.; Seidman, C.; Ljunggren, S.; Lefeber, D.; Morava, E.; Wevers, R.; Fritz, T.; Tabak, L.; Lindahl, M.; Hovingh, G.; Kuivenhoven, J. Heterozygosity for a Loss-of-Function Mutation in GALNT2 Improves Plasma Triglyceride Clearance in Man. *Cell Metabolism* 2011, 14 (6), 811-818.





# Chapter 7

Ultrahigh resolution profiles lead to more  
detailed serum peptidome signatures  
of pancreatic cancer

*Simone Nicolardi, Berit Velstra,  
Bart J. Mertens, Bert Bonsing  
Wilma E. Mesker, Rob A.E.M. Tollenaar,  
André M. Deelder, and Yuri E.M van der Burgt*

Translational Proteomics

2014, 2, 39-51

## **ABSTRACT**

Mass spectrometry-based (clinical) proteomics has been widely applied as a technology to find and validate disease-specific protein signatures. MALDI-based peptidome profiles provide a suitable platform for classification of body fluids or tissues, albeit at the cost of being unable to observe low abundant species. Here we show that a fully automated one-step solid-phase extraction serum sample cleanup in combination with fast MALDI acquisition and ultrahigh precision 15 tesla FTICR readout provides a powerful, fast and robust approach for obtaining biomarker signatures. This is exemplified for a cohort of pancreatic cancer patients. Specific “early cancer” symptoms such as pain, jaundice or weight loss are often not experienced, thus delaying diagnosis of the disease. Novel markers for early diagnosis of pancreatic cancer are therefore urgently needed. A total of 273 serum samples, distributed over a calibration and validation set, was processed and mass analyzed within a time frame of 24 hours. In both sets sensitivity and selectivity values were well above 85%. In these “next-generation” MALDI peptidome profiles all species up to 9 kDa were isotopically resolved. Finally, it is noted that the low ppm mass accuracy of peptides and proteins observed between 1 and 9 kDa in the FTICR profiles facilitates sequence identifications.

## INTRODUCTION

Pancreatic cancer (PC) is the fourth (females) and fifth (males) leading cause of cancer death in developed countries, with a relatively low annual incidence of 5.4 cases per 100,000 females and 8.2 cases per 100,000 males [1]. Patients often die within the first half year after diagnosis, or have an extremely poor prognosis with an overall five-year survival rate of less than 5% [2]. When surgical resection is possible, five-year survival rates improve to approximately 25%. Unfortunately, when the first symptoms appear most tumors are at an advanced stage and their surgical resection would not improve the prognosis [3;4]. Molecular biomarkers that detect PC at an early stage with high sensitivity and specificity would thus be highly beneficial. At the moment, the only used blood marker for detecting and following PC in the clinic is the mucin-associated carbohydrate antigen CA 19-9. This marker, however, often fails in detecting small, resectable cancers [5]. Consequently, like in other cancer biomarker studies, serum proteomics has become a popular approach to find new markers for PC, since blood is a rich and powerful source of biomarkers in general and samples can be collected in a minimally invasive way. The discovery of serum biomarkers is mainly performed by mass spectrometry (MS)-based proteomics methods [6]. One of these involves the comparison of serum protein profiles in a “case versus control” manner by matrix-assisted laser desorption/ionization - time of flight (MALDI-TOF) MS [7]. Such profiles (*i.e.* mass spectra) contain hundreds of features (or peaks), of which the presence and intensity can depend on the physiological and pathological condition of the individual. The statistical analysis of serum peptide and protein profiles obtained from both control and diseased individuals allows the identification of a set of features, or a so-called biomarker signature, that can be valuable in understanding the specific disease. Moreover, the biomarker signature may provide leads to further exploit diagnostic and therapeutic potential. Encouraging results have been obtained using profiling strategies [8-10]. Nevertheless, the route to clinically applicable protein assays faces various types of challenges [11;12]. With regard to the selected methodology, for MS-based peptide profiling approaches the problems can be categorized as follows. First of all, multiple profiling studies have shown to lack reproducibility and could not be validated. In this context, standardization of the protocols used for serum sample collection and for peptide and protein purification is pivotal [10;13;14]. The use of a fully automated high-



throughput platform for sample processing based on solid-phase extraction (SPE) has been shown to minimize variation and to improve robustness of the method [15]. Secondly, previous MS-acquisitions such as performed on surface-enhanced laser desorption/ionization (SELDI) platforms were not robust and yielded poor accuracies. In addition, identification of peptides or proteins was cumbersome, or not possible at all in these early profiling studies. However, with current equipment these issues can be considered obsolete. The use of internal standards in combination with modern mass analyzers now allows precise quantitation and detailed characterization of peptides in high-throughput profiles [16;17]. Thirdly, similar peptide profiles were found for various diseases, implying that the features were not specific. On the other hand, it has been postulated that well-defined degradation of highly abundant proteins into peptides (“degradome”) can result in tumor-specific serum peptidome patterns [18].

Recently, we reported a protein profiling study for PC performed on a fully automated SPE-based serum processing platform [19]. Proteins were first isolated with weak cation exchange (WCX) magnetic beads (MBs) using a 96-channel liquid handling robot, followed by acquisition of linear mode MALDI-TOF profiles in the range of 1 to 12 kDa, and evaluation via linear discriminant analysis with double cross-validation. This resulted in a discriminating WCX-profile for PC with a sensitivity of 78% and a specificity of 89% in the calibration set with an area under the curve (AUC) of 90%. These results were validated with a sensitivity of 74% and a specificity of 91% (AUC 90%). However, an obvious disadvantage of low resolution MS profiles is the fact that (poly)peptides and proteins are measured as broad peaks, thus leading to one of the earlier mentioned problems on peak identification. In a second profiling study using the same PC cohort, serum samples were processed with reversed-phase (RP) C18 MBs, and resulting peptides were measured with high-resolution reflectron mode MALDI-TOF MS yielding isotopically resolved profiles up to 4 kDa. For statistical evaluation, a list of 42 different peptides was compiled from which a discriminating profile for PC could be defined, with an area under the curve (AUC) of 92% (98%) a sensitivity of 76% (95%) and specificity of 91% (100%) in the calibration (validation) set. Although the identity of most of these peptides was known or elucidated, it became clear that multiple peptides still overlapped at the resolving power of approximately 11,000 [20]. The effect of increased resolving power was therefore further studied in MALDI-profiles obtained by Fourier transform ion

cyclotron resonance (FTICR) MS, a platform that has proven to be extremely powerful for the analysis of complex mixtures, such as oil, organic matter and plasma [21-23]. With proper control, mass resolving powers higher than 100,000 (at  $m/z$ -value 1,000 with 1s transient) and low or sub-ppm mass measurement errors can be routinely obtained [24;25]. We have previously developed a MALDI-FTICR workflow on a commercially available platform equipped with a 15 tesla magnet that allows high-throughput and fully automated profiling of human serum peptides and proteins with isotopic resolution up to 15,000 Da [26;27]. By following this approach, in comparison to high resolution TOF analyzers the spectrum alignment is more accurate and the quantification of peptides more robust due to the improved mass measurement precision. In this study this MALDI-FTICR workflow in combination with SPE-based sample cleanup with RPC18-functionalized MBs was applied for the analysis of a clinical cohort. Here, “next-generation” MALDI-FTICR peptide and protein profiles were generated using serum samples obtained from PC patients and control individuals (258 samples in total). Classification performances of both the calibration and validation set were compared to those previously obtained from the same PC cohort, either processed with different MBs or measured on a different mass analyzer. Discriminating peaks (*i.e.* a biomarker signature) defined from the calibration set were validated using an independent case-control group. Finally, the low ppm mass accuracy provided by the MALDI-FTICR platform narrows the search window for *de novo* identifications of peptides and proteins in the profiles.

## MATERIALS AND METHODS

**Patients and sample collection.** For the calibration set, serum samples were obtained from 49 patients with PC prior to surgery, and from 110 (age- and gender-matched) healthy volunteers (“controls”) over a time period ranging from October 2002 until December 2008 at the outpatient clinic of the Leiden University Medical Center (LUMC), the Netherlands. Healthy volunteers were partners or accompanying persons of included patients. For the validation set, serum samples were obtained from 39 patients and 75 healthy (age- and gender-matched) volunteers over a time period ranging from January 2009 until July 2010. Patients were selected candidates for curative surgery, thus no patients with primary irresectable tumours were included. All surgical specimens were

examined according to routine histological evaluation and the extent of the tumor spread was assessed by TNM (TNM Classification of Malignant Tumors) classification. Informed consent was obtained from all subjects and the study was approved by the Medical Ethical Committee of the LUMC. All samples were collected and processed according to a previously reported standardized protocol [9]. Briefly, blood samples were drawn by antecubital venipuncture while the individuals, who had not been fasting prior to any invasive procedure, were seated. The samples were collected in an 8.5 mL Serum Separator Vacutainer Tube (BD Diagnostics, Plymouth, UK) and maximally within 4 h at room temperature were centrifuged at 1000 g for 10 min. Serum samples were then distributed into sterile 500- $\mu$ L barcode labelled polypropylene aliquots (TrakMate; Matrix TechCorp.) and stored at  $-80^{\circ}\text{C}$ . All serum samples were thawed on ice once and randomly placed in barcode labelled racks in an 8-channel Hamilton STAR<sup>®</sup> pipetting robot (Hamilton) for automated aliquotting into 60- $\mu$ L daughter tubes. The aliquots were stored in 96-tubes racks at  $-80^{\circ}\text{C}$  until further sample processing. Samples from the calibration and the validation set were distributed over three 96-tubes racks as following: one full 96-tube rack for both the calibration and validation set and one partially-filled 96-tube rack with 63 samples from the calibration set and 18 samples from the validation set. Identical processing steps were followed for the two sample sets.

**High-throughput RPC18-MB chromatography and MALDI spotting.** The isolation of peptides from human serum was performed using RPC18-functionalized MBs as previously described [27]. In short, RPC18-MBs were first activated by a three-step washing with a 0.1% TFA solution. Then, for each sample 5  $\mu$ L of serum was added to the activated beads and incubated for 5 min at room temperature. The beads were washed again three times with 0.1% TFA and peptides were eluted with a 1:1 mixture of water and acetonitrile. Two microliters of each (stabilized) eluate were mixed with 10  $\mu$ L of an  $\alpha$ -cyano-4-hydroxycinnamic acid MALDI matrix solution in a 384-well PCR plate. Then, 1  $\mu$ L of this mixture was spotted in quadruplicate onto a 600  $\mu\text{m}$  Anchor-Chip<sup>™</sup> MALDI-target plate (Bruker Daltonics). The so-called RPC18 eluates from the calibration and the validation set were spotted onto three 384-spots MALDI-target plate as following: 96 eluates from the calibration set and 96 eluates from the validation set were spotted in quadruplicate onto two distinct MALDI-target plates; the remaining eluates from the two

sets were spotted in quadruplicate onto the same MALDI-target plate. This SPE- and MALDI-spotting procedure requires approximately 3 hours per plate of 96 samples.

**MALDI-FTICR mass spectrometry and data processing.** MALDI-FTICR experiments were performed on a Bruker 15 tesla solariX™ FTICR mass spectrometer equipped with a novel CombiSource (Bruker Daltonics). The MALDI-FTICR system was controlled by Compass solariXcontrol software and equipped with a Bruker Smartbeam-II™ laser system that operated at a frequency of 200 Hz. The ‘medium’ predefined shot pattern was used for the irradiation. Two acquisition settings, namely low-mass method (LM) and high-mass method (HM), respectively, were used to optimize both the sensitivity and resolving power in the mass range from 1013 to 3700 Da and in the mass range from 3500 to 10000 Da, respectively. These methods were optimized as previously described with some modification [27]. For both methods, each mass spectrum was obtained from the sum of 10 scans of 150 laser shots each and using 512 K data points. Typically, the target plate offset was 100 V with the deflector plate set at 180 V. The ion funnels operated at 100 V and 6.0 V, respectively, with the skimmers at 15 and 5 V. The analyzer entrance was maintained at -7 V, and side kick technology was used to further optimize peak shape and signal intensity. The trapping potentials were set at 0.60 and 0.55 V for the LM and at 0.95 and 0.80 for the HM. The required excitation power was 28% with a pulse time of 20.0  $\mu$ s. The two acquisition settings differentiate for the trapping potentials (LM, 0.6 and 0.55 V; HM, 0.95 and 0.80 V), the required excitation power (LM, 25%; HM, 28%) and pulse time (LM, 10  $\mu$ s; HM, 20  $\mu$ s), the time of flight to the ICR cell (LM, 1.350 ms; HM, 2.700 ms) and the quadrupole filter mass (LM,  $m/z$  1300; HM,  $m/z$  2500). For each spotted sample, two duplicate spots were measured using the LM and the other two using the HM. Approximately 4.5 hours were needed to measure 384 MALDI spots (*i.e.* originating from 96 different serum samples). DataAnalysis Software 4.0 SP 5 (Bruker Daltonics) was used for the visualization and the calibration of the spectra. Prior to the measurement of each MALDI plate the FTICR system was externally calibrated using a commercially available peptide mix and a protein mix (Bruker Daltonics). The spectra obtained using the LM were internally calibrated only when used for identification purposes. The  $m/z$ -values used for the internal calibration of the LM and the HM are reported in Table S1 in the Supplementary Material section. Peaks were

determined using the FTMS algorithm with a signal-to-noise threshold of 3 and using the centroid for peak position with a percentage height of 80.

**Peak selection and quantification.** Protein and/or peptide signals in RPC18 profiles were quantified as follows. First, based on visual inspection of the profiles, 457 and 670 peaks were selected for the LM and HM spectra, respectively, for further analysis. To this end, a so-called reference file was compiled for both types of profiles in such a way that for each selected peak the  $m/z$ -value, a peak number and an  $m/z$ -window were reported. In the LM profiles, this  $m/z$ -window ranged from 0.015 to 0.166 Da while in the HM it ranged from 0.05 to 0.31 Da reflecting the peak width along the spectra. Then, the in-house developed Xtractor tool was used to determine the intensity of each user-defined peak. This open source tool generates uniform data (peak) arrays regardless of spectral content (<http://www.msutils.org/Xtractor>). MALDI-FTICR profiles were exported as XY (.xy) files, all containing  $m/z$  values with corresponding intensities. Although peptide and proteins were measured up to 10,000 Da using the HM method, the peak selection was limited to 9043.3 Da. The analysis of the spectra in the  $m/z$ -range from 9043.3 to 10000 is on-going and the results will be presented in a separate study.

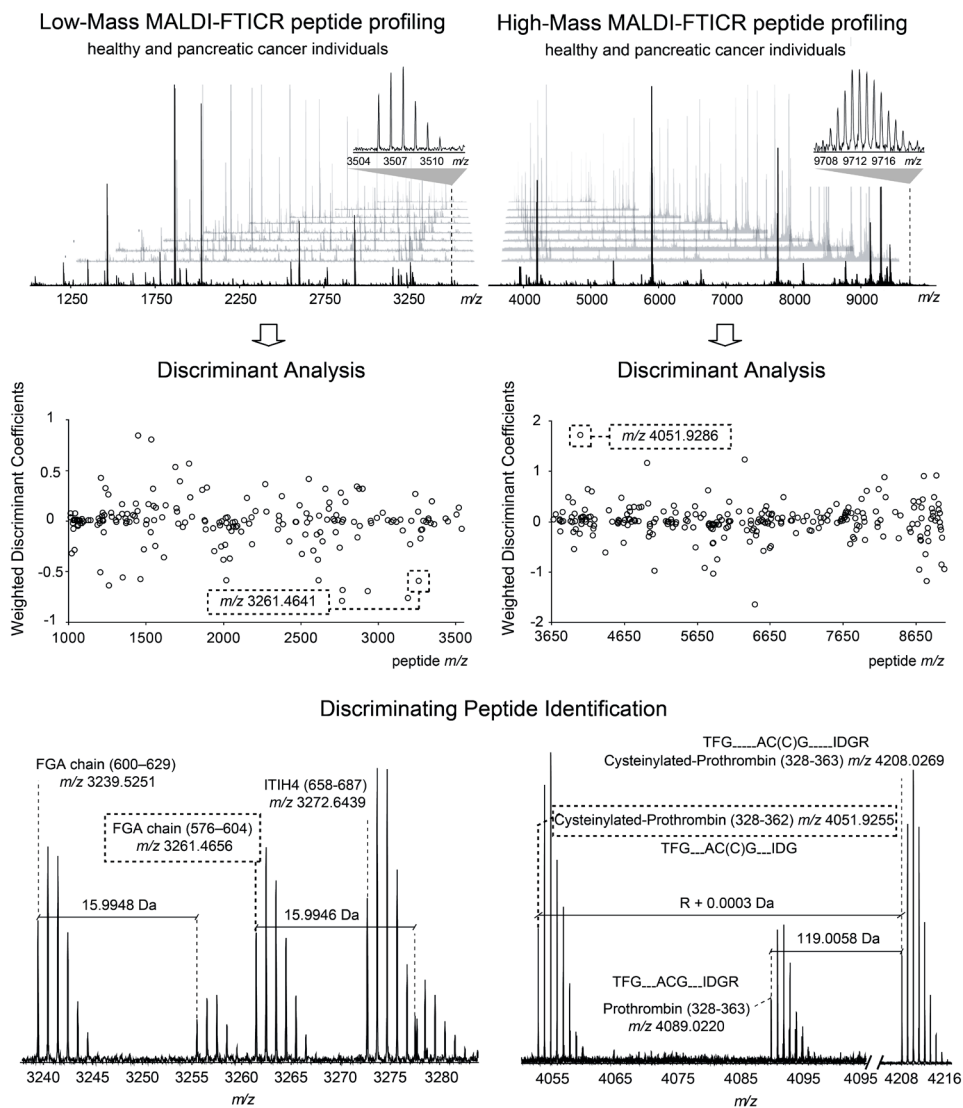
**Statistical analysis.** Peak intensities were transformed using the logarithmic function, followed by calculation of the mean of peak intensities obtained from replicate spectra. The intensities of isotope peaks belonging to the same peptide were further summed to reduce the number of features and time needed for further analysis. For each sample, 196 and 291 peak intensity values were obtained for the LM and HM, respectively, and were used for statistical analysis. To this end, logistic regression ridge shrinkage (LRRS) analysis was applied to the calibration sets (*i.e.* LM and HM data from the calibration set) in order to calibrate two diagnostic rules for the classification of the serum sample either as case or control. Each sample was assigned to the group for which the probability was higher. The prediction rules obtained from the application of LRRS on the calibration sets were applied to the validation sets (*i.e.* LM and HM data from the validation set). Thus, each sample was classified and the results were compared with known disease status. The classification probabilities assigned to each sample using the LM and HM data from the validation set were further combined. To this end, LRRS analysis was performed

on the combination of the Logit transformed probabilities obtained for validation sets. This analysis involves the recalibration of the validated diagnostic rule. For each analysis error rate (error = the amount by which an observation differs from its expected value), sensitivity, specificity and area under the curve (AUC) were calculated. The error rates are based on the sensitivity and specificity values, assuming a prior class probability of 0.5 for each group. Receiver-operating characteristic (ROC) curves with the true-positive rate (sensitivity) were plotted in function of the false-positive rate (1-specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between groups (diseased/healthy). Univariate discriminate analysis was performed to determine which peak varied the most between case and control groups. This study was limited to peaks of which the absolute weighted discriminant coefficient was higher than 0.1 in the multivariate discriminant analysis used to calibrate the discriminant models. Finally, a t-test was performed on a selection of peaks for the calibration sets only.

## RESULTS

**MALDI-FTICR-MS peptide and protein profiling.** Serum samples of PC patients as well as control individuals were processed simultaneously using a previously described fully automated and standardized SPE-based RPC18-MB protocol [15]. Thus obtained MB eluates were spotted onto a MALDI target plate in quadruplicate. Two types of ultrahigh resolution peptide and protein profiles were then acquired applying an automated acquisition procedure on the MALDI-FTICR system (see Materials and Methods section). Two out of four spots were used to obtain a so-called low mass (LM) profile ( $m/z$ -values from 1013 to 3700) and the remaining two spots were used to generate a so-called high mass (HM) profile ( $m/z$ -values from 3500 to 10,00). In total, 273 serum samples were analyzed in this way, thus yielding 1,092 profiles. A typical example of both an LM and HM MALDI-FTICR profile is depicted in Figure 1 (upper panel). It was verified that all peptides and (small) proteins were measured with isotopic resolution through all the spectra, with typical resolving powers varying from 130000 ( $m/z$  1039.6727) to 46000 ( $m/z$  3523.7664) in the LM spectra and from 150000 ( $m/z$  3680.8709) to 33000 ( $m/z$  9744.6054) in the HM spectra (as plotted in Figure 2A).

As a result, a large number of peptides or proteins that would overlap in high resolution MALDI-TOF MS were measured as distinct features by MALDI-FTICR MS. Two examples of resolved species are shown in Figure 2B, one for the LM and one for the HM profiles. The ultrahigh resolving power allowed the accurate quantification of the selected peptides and proteins in all the spectra. After manual inspection of the profiles, 457 and 670 peaks for the LM and HM, respectively, were selected for statistical analysis. After taking into account isotopic peaks from the same species, 196 peptides remained from the 457 selected peaks in LM spectra and 291 peptides or proteins remained from the 670 selected peaks in HM spectra. Peptides and proteins were detected with signal intensities that typically ranged over two orders of magnitude. For example, Fibrinopeptide alpha chain (2–16) (at  $m/z$ -value 1465.6554) was often observed as the most intense peptide, and was 304 times more intense than Complement C4-A (1337–1350) (at  $m/z$ -value 1626.8459) detected with a signal-to-noise ratio (S/N) of 6.6, in a typical spectrum. Thus, peptides observed with low S/N were also evaluated. For example, the peptide identified as oxidized Fibrinogen beta chain (45-71)( $m/z$  2898.5334) (see Serum peptide identification by accurate mass difference measurement section) was observed in the spectra in the calibration set with an averaged S/N 9.6 with a standard deviation (SD) of 6.4, while the highly intense Complement C3f fragment peptide (at  $m/z$ -value 2021.1039) was observed with an averaged S/N of 2035 with an SD of 345. As a final remark, from 12 out of 1,032 profiles the quality was insufficient for further statistical analysis, most likely because of failed MALDI spotting.

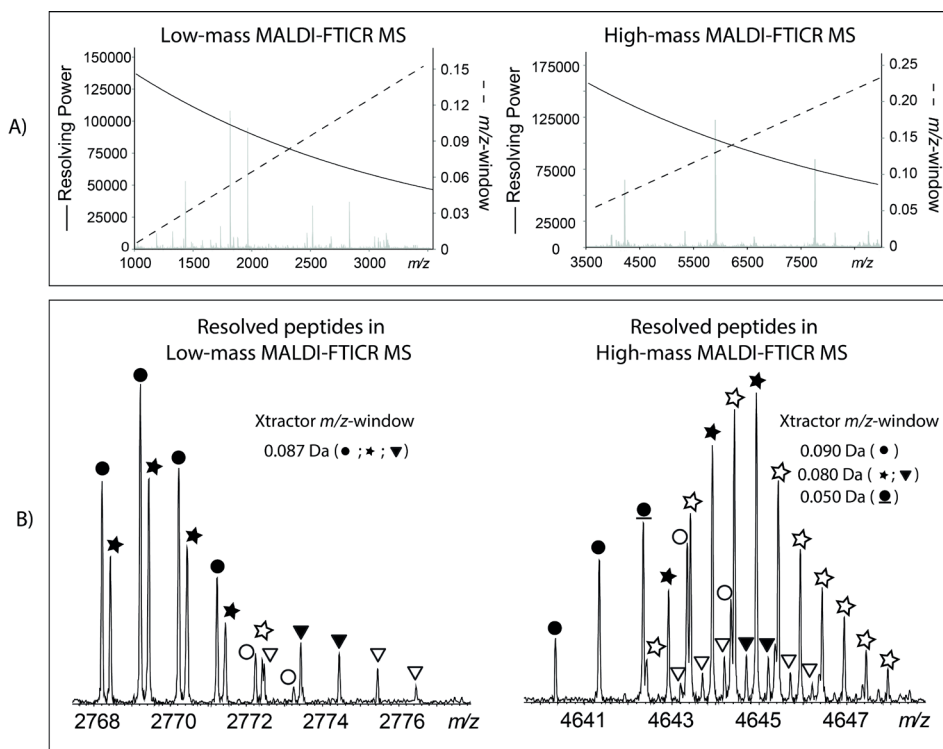


**Figure 1.** Ultrahigh resolution 15T MALDI-FTICR peptide and protein profiles obtained from human serum after RPC18 magnetic bead-based SPE (upper panel). Using two different acquisition settings, optimized for the low-mass and high-mass range, peptides and proteins were isotopically resolved up to 10 kDa. For statistical analysis, logistic regression ridge shrinkage analysis was used to find changes in the peptide and protein profiles obtained from



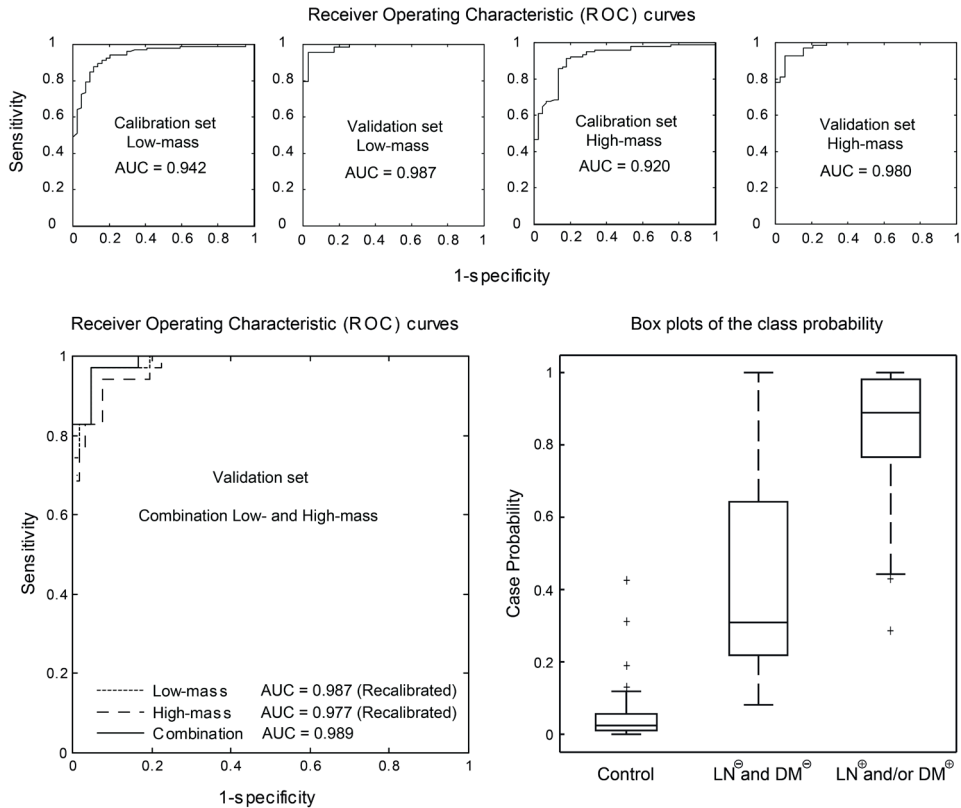
healthy individuals and pancreatic cancer patients. A weighted discriminant coefficient was assigned to each detected peptide or protein according to its discriminant property in the way that the higher the value of the discriminant coefficient the higher the case probability (middle panel). Identifications of the most discriminating peptides and proteins were based on previously reported peptide IDs or on accurate mass measurement of mass differences in the spectra (lower panel). Note that these identifications need further confirmation by MS/MS-data.

**Statistical analysis.** The signal intensities of all selected peaks were determined in all serum profiles using the Xtractor tool described in the Materials and methods section. As shown in Figure 2A, the  $m/z$ -windows in the reference files were fine-tuned according to the resolving power calculated for each  $m/z$ -value. The presence of different peptides with close masses was also taken into account as well as the mass measurement precision (see Figure 2B). The optimization of this  $m/z$ -window allowed the accurate quantification of all peaks selected from the spectra. Thus obtained peak intensity values were then used for statistical analysis. To this end, a discriminate model was first calibrated and then validated using LRRS analysis on the calibration and validation sets, respectively. The ROC curves resulted from this analysis are shown in Figure 3. Error rates of 0.136 and 0.104, sensitivities of 88% and 91% and specificities of 96% and 94% with AUC of 0.987 and 0.980 were obtained for the LM and HM validation sets, respectively. The LRRS analysis performed on the combination of the logit of the classification probabilities obtained for the LM and HM validation sets resulted in an error of 0.0784, a sensitivity of 89% and a specificity of 100% with an AUC of 0.989.



**Figure 2.** For each isotopically resolved peptide or protein signal a specific  $m/z$ -window was defined to allow accurate quantification. The  $m/z$ -window was based on the resolving power at a specific  $m/z$ -value (A) and in the case of overlapping species further optimized (B). Note that the resolving power at  $m/z$  3500 is higher in the High-mass measurements (*i.e.* ~ 150000) than in the Low-mass measurements (*i.e.* ~ 46000) as a result of the specific broadband frequency-sweep waveform that was used to excite the ions into the ICR cell.

The logit transformation involves a recalibration of the discriminant models obtained using the validation sets. The discriminant analysis performed on the recalibrated validation sets resulted in errors of 0.098 and 0.088, sensitivities of 88% and 90% and specificities of 96% and 93% with AUC of 0.987 and 0.977 for the LM and HM validation sets, respectively.



**Figure 3. Receiver Operating Characteristic (ROC) curves generated from the case-control classification of the serum samples from both the calibration sets and the validation sets using Low-mass (LM) and High-mass (HM) data (four upper panels). The larger the area under the ROC curve (AUC), the better two groups (diseased/healthy) can be distinguished. The lower panel on the left hand-side shows ROC curves generated from the combination of the classification results obtained from the validation set. This combination involved the recalibration of the validated diagnostic rules, resulting in recalibrated LM and HM data. The lower panel on the right hand-side shows box plots that are obtained after sub-typing the PC cases into cases without any metastasis (*i.e.* regional lymph node-negatives (LN<sup>-</sup>) and no distant metastasis (DM<sup>-</sup>)) versus lymph node-positives (LN<sup>+</sup>) and/or distant metastasis (DM<sup>+</sup>) based on TNM-classification.**

A sequential analysis was performed by sub-typing the PC cases into cases without any metastasis (*i.e.* regional lymph node-negative (LN-) and no distant metastasis (DM-)) versus cases that were lymph node-positives (LN+) and/or showed distant metastasis (DM+), based on TNM-classification summarized in Table 1. This sub-typing resulted in a box plot (see Figure 3) with separation between controls and cases, and in addition good separation between cases with and without metastasis (Wilcoxon Mann-Whitney test with a p-value of 7.7293e-05 for controls versus “(LN-)and(DM-)”, and a p-value of 0.015844 for “(LN+)and/or(DM+)” versus “(LN-)and(DM-)”). Patient characteristics, number of serum samples, and the results of the classification methods set are shown in Table 1.

**Table 1. Patient characteristics for the calibration and validation sets.**

	Calibration Set			Validation Set		
	HM [Correct Classif.]	LM [Correct Classif.]	Combination HM and LM	HM [Correct Classif.]	LM [Correct Classif.]	Combination HM and LM
N. of samples	46	45		39	35	
Age median (min - max)	65.6 (41-80)	65.2 (41-80)		63.4 (38-81)	62.8 (38-81)	
Male	23 [16]	22 [17]	16	17 [15]	15 [11]	12
Female	23 [15]	23 [15]	17	22 [17]	20 [15]	15
<b>Localisation</b> (pancreas)						
Head	38 [28]	36 [27]	33	35 [28]	31 [22]	23
Body	3 [1]	3 [2]	1	1 [1]	1 [1]	1
Tail	3 [1]	4 [2]	2	1 [1]	1 [1]	1
Unknown/Other	2 [1]	2 [1]	1	2 [2]	2 [2]	2
<b>Stage</b>						
IA	7 [7]	7 [6]	7	2 [1]	2 [0]	0
IB	5 [3]	5 [3]	3	2 [2]	1 [1]	1
IIA	3 [2]	3 [1]	2	3 [1]	3 [0]	1
IIB	19 [13]	17 [14]	13	21 [18]	20 [17]	17
III	5 [3]	5 [4]	3	2 [2]	2 [2]	2
IV	7 [3]	8 [4]	5	9 [8]	7 [6]	6
<b>Tumor differentiation</b>						
Unknown (irresectable tumor)	12 [7]	11 [6]	7	11 [10]	10 [9]	9
Grade 1	10 [5]	11 [5]	6	6 [2]	6 [3]	3
Grade 2	11 [9]	11 [9]	9	15 [14]	13 [10]	10
Grade 3	13 [10]	12 [12]	11	7 [6]	6 [4]	5

A logistic regression coefficient weighted by the standard deviation of the peak intensity was assigned to each peak as determined from multivariate analysis on the calibration set (*i.e.* the calibration of the discriminating rule). These discriminant weights denote the conditional effect associated with each peak, after taking into account the variation in expression across the other selected peaks. Thus, the higher the value of the discriminant weight the higher the case probability. Note that the reverse applies to control samples. The plots with the weighted discriminant coefficients *vs* the *m/z*-values are shown in Figure 1 (middle panel). A t-test was performed on peaks with absolute discriminant coefficient higher than 0.1 in the calibration set. A P-value smaller than 0.001 was considered as significant. Peaks that satisfied these criteria are reported in Table 2 with corresponding protein names, t-test values, standard deviations (SD), P-values, 95%-confidence interval and the weighted discriminant coefficients. Note that the P-values here reported ranged from  $6.0 \times 10^{-4}$  to  $4.0 \times 10^{-9}$  indicating a high statistical significance.

**Table 2. Peptides present in MALDI-FTICR precision profiles with low P-values after comparative analysis of serum samples from pancreatic cancer patients and control individuals.**

Observed <i>m/z</i> [M+H] <sup>+</sup>	Protein Name	T-test	SD	Pvalue	CI lower	CI upper	Bweighted
Low-Mass MALDI-FTICR profiles							
1206.5753	FPA chain (5-16)	3.7	3.4	2.7E-04	1.1	3.5	-0.51
1211.6535	Complement C3f fragment (7-16)	-4.0	2.0	1.1E-04	-2.1	-0.7	0.42
1263.5965	FPA chain (4-16)	5.0	3.0	1.3E-06	1.6	3.7	-0.64
1348.7118	Complement C3f fragment (6-16)	-3.6	1.5	4.1E-04	-1.5	-0.4	0.16
1350.6282	FPA chain (3-16)	4.4	3.5	2.4E-05	1.5	3.9	-0.56
1449.7597	Complement C3f fragment (5-16)	-4.0	3.3	1.1E-04	-3.5	-1.2	0.84
1561.7265	Protrombin (315-327)	4.5	1.9	1.4E-05	0.8	2.2	-0.25
1616.6584	FPA chain 1P	5.0	3.4	1.8E-06	1.8	4.2	-0.36
1626.8458	Complement C4-A (1337-1350)	-3.7	1.1	3.2E-04	-1.1	-0.3	0.15
1698.7450	n.i.	-3.5	0.9	6.0E-04	-0.9	-0.3	0.12
1718.9430	Complement C3f fragment (4-16)	-4.2	1.6	4.4E-05	-1.7	-0.6	0.39
1786.8545	ITIH4 (671-687)	-4.7	1.7	6.5E-06	-2.0	-0.8	0.23
1984.9879	n.i.	3.6	1.2	4.8E-04	0.3	1.2	-0.25
2685.3746	Thrombin light chain (342-363)	-5.8	1.1	3.9E-08	-1.5	-0.7	0.18
2768.2293	FGA chain (576-600)	3.9	3.5	1.5E-04	1.2	3.7	-0.79
2898.5334	FGB chain (45-71) + 15.9952 Da	-6.3	1.3	4.0E-09	-2.0	-1.0	0.32
2931.2909	FGA chain precursor (576-601)	4.9	3.8	2.2E-06	2.0	4.7	-0.70
3190.4279	FGA chain (576-603)	5.9	3.1	2.9E-08	2.1	4.3	-0.76
3206.4243	FGA chain (576-603) [Met-ox]	5.2	2.1	7.1E-07	1.2	2.7	-0.23
3261.4664	FGA chain (576-604)	5.1	3.5	1.3E-06	1.9	4.4	-0.59
High-Mass MALDI-FTICR profiles							
3679.8665	n.i.	-4.3	0.7	3.7E-05	-0.8	-0.3	0.14
3806.8902	n.i.	-4.2	0.7	5.4E-05	-0.7	-0.3	0.19
4051.9297	Thrombin light chain (328-363) (Cysteinylated)	-5.8	2.8	4.4E-08	-3.8	-1.9	1.72
4108.9479	Thrombin light chain (328-363) (Cysteinylated+Gly)	-4.6	1.5	7.6E-06	-1.8	-0.7	0.24
4394.0803	n.i.	3.8	1.6	2.1E-04	0.5	1.7	-0.62
4854.2768	Apolipoprotein CIII2 (21-99) *	-3.9	2.5	1.7E-04	-2.6	-0.8	0.28
4961.4906	FGA chain (529-574) or (513-558)	-4.5	2.9	1.3E-05	-3.3	-1.3	1.16
4979.4945	FGA chain (529-574) or (513-558) + 15.9952 Da	-5.0	2.1	1.7E-06	-2.6	-1.1	0.36
4985.4806	n.i.	-3.7	1.7	2.9E-04	-1.7	-0.5	-0.30
5802.6370	FGA chain (576-627)	-3.7	3.2	2.7E-04	-3.2	-1.0	0.62
6223.3096	n.i.	-4.5	1.4	1.2E-05	-1.6	-0.6	0.19
7151.4191	n.i.	-3.5	1.6	5.9E-04	-1.5	-0.4	0.11
8205.0930	Apolipoprotein CII (23-101)	-3.8	1.8	2.4E-04	-1.9	-0.6	0.88
8781.2819	Apolipoprotein-CIII (21-99) + 15.9952 Da	3.8	2.8	1.8E-04	0.9	2.9	-1.19

\* This species was observed as a doubly-charged ion.

**Serum peptide identification by accurate mass difference measurement.** A list of serum peptides and proteins that are commonly observed in MALDI-TOF profiles obtained after RPC18-based sample cleanup has been compiled previously and this was used for statistical evaluations [9]. However, a number of peptides remained unidentified in this list, and moreover in the current MALDI-FTICR ultrahigh resolution profiles many RPC18-MB serum eluate peaks are unknown. Likely, a large number of these degradome peptides originate from the same high abundant proteins after proteolytic cleavage as was reported earlier [18;28;29]. New peptide assignments were performed based on matching accurate mass measurements of  $m/z$ -differences between peaks in 15T MALDI-FTICR spectra with possible decreased or increased sequences (“degradome”). Thus, a search for consecutive mass differences corresponding to one amino acid was performed, starting from a previously identified peptide in the spectrum with relatively highest signal intensity. In this way, new peptides with one or more additional amino acids at the N-terminus or/and the C-terminus or modified peptides (*i.e.* oxidized, cysteinylated) were identified. Following this strategy the amino acid sequence of 34 new peptides was derived and these are reported in Table 3. In general, the LM and HM profiles provided sub- and low-ppm mass measurement errors for these identifications, respectively. Two examples of this approach are shown in Figure 1 (lower panel). The first one is the identification of an oxidized form of the peptide Fibrinogen alpha chain (576–604) that was statistically evaluated with a discriminant weight factor of -0.59 (see Table 2). In the second example the accurate mass-based identification of the species observed at  $m/z$ -value 4051.9255 is depicted, a peptide that was found to be the best predictor (*i.e.* highest absolute discriminant weight) of healthy and disease individuals in HM profiles (see Table 2). The mass difference between this peptide and a peptide previously MS/MS-identified as cysteinylated-Prothrombin (328-363), observed at  $m/z$ -value 4208.0269, was 156.1014 Da. This mass difference corresponds to an arginine residue with an error of only 0.3 milliDa. In addition, the accurate measurement of mass differences allowed the identification of peptides containing a single amino acid mutation. For example, a peptide from coagulation factor XIII (Factor XIIIa) alpha chain with a previously reported Val35Leu mutation corresponding to a mass difference of 14.0156 Da between “normal” and mutant fragment peptides was indeed observed (see Table 3). Here, the species at  $m/z$ -value 2602.3113 corresponds to a previously identified peptide from Factor XIIIa (14-38),

whereas the species at  $m/z$ -value 2531.2735 and  $m/z$ -value 2545.2883 both lack an alanine residue but differ at the site of mutation (*i.e.* Val35 Factor XIIIa (15-38) and Leu35 Factor XIIIa (15-38), respectively). It is emphasized that isobaric peptides containing modifications such as oxidation cannot be uniquely characterized by the accurate measurement of mass differences. For this purpose additional MS/MS-experiments are needed to confirm the identifications and localize for instance modified amino acids in the sequence. As a final remark, the accurate and precise MALDI-FTICR mass measurements will allow a reliable match between the MS/MS-data obtained using other MS techniques such as LC-ESI-MS/MS and the peptides observed in the MALDI-FTICR spectra.



**Table 3. Proposed sequences of (large) peptides present in MALDI-FTICR precision profiles based on accurate mass measurements.**

Protein Name	Peptide Sequence	Mass difference	Calculated $m/z$ [M+H] <sup>+</sup>	Observed $m/z$ [M+H] <sup>+</sup>	Mass Measur. Error (ppm)	MALDI-FTICR method
ITIH4 UniProt Q14624	M.NFRPGVLSSRQLGLPGPPDVP DHAAYHPF.R	+(NRF + 0.0033)	3141.6017	3141.6047	0.96	LM
	N.FRPGVLSSRQLGLPGPPDVPD HAAAYHPF.R	+(RF + 0.0009)	3027.5588	3027.5593	0.18	LM
	R.PGVLSRQLGLPGPPDVPDHA AYHPF.R or RPGVLSSRQLGLPGPPDVPDHA AYHPF.R	+(R - 0.0004)	2880.4904	2880.4896	-0.26	LM
	<b>R.PGVLSRQLGLPGPPDVPDH AAYHPF.R</b>	-	<b>2724.3893</b>	<b>2724.3889</b>	<b>-0.13</b>	<b>LM</b>
	P.GVLSRQLGLPGPPDVPDHA YHPF.R	-(P + 0.0002)	2627.3365	2627.3360	-0.19	LM
Thrombin light chain UniProt P00734	L.FEKKSLEDKTERELLESYIDGR	+(FEKK - 0.0011)	2685.3730	2685.3720	-0.36	LM
	F.EKKSLEDKTERELLESYIDGR	+(EKK - 0.0008)	2538.3046	2538.3039	-0.28	LM
	E.KKSLEDKTERELLESYIDGR	+(KK - 0.0005)	2409.2620	2409.2615	-0.19	LM
	K.KSLEDKTERELLESYIDGR	+(K + 0.0004)	2281.1670	2281.1676	0.23	LM
	<b>K.SLEDKTERELLESYIDGR</b>	-	<b>2153.0721</b>	<b>2153.0722</b>	<b>0.04</b>	<b>LM</b>
	S.LEDKTERELLESYIDGR	-(S + 0.0012)	2066.0400	2066.0390	-0.53	LM
	L.EDKTERELLESYIDGR	-(SL + 0.0001)	1952.9560	1952.9560	0.02	LM
	E.DKTERELLESYIDGR	-(SLE + 0.0010)	1823.9134	1823.9124	-0.52	LM
D.KTERELLESYIDGR	-(SLED + 0.0001)	1708.8864	1708.8865	0.01	LM	
Thrombin light chain UniProt P00734	TFGSGEADC(CysGly)GLRPLFEK KSLEDKTERELLESYIDG	+(G + 0.0032)	4265.0489	4265.0520	0.73	HM
	<b>TFGSGEADC(Cys)GLRPLFEKK SLEDKTERELLESYIDGR</b>	-	<b>4208.0275</b>	<b>4208.0274</b>	<b>-0.02</b>	<b>HM</b>
	TFGSGEADC(CysGly)GLRPLFEK KSLEDKTERELLESYIDG	-(R-Gly-0.0001)	4108.9478	4108.9479	0.02	HM
	TFGSGEADCGLRPLFEKKSLEDK TERELLESYIDGR	-(Cysteinylation + 0.0010)	4089.0234	4089.0222	-0.27	HM
	TFGSGEADC(Cys)GLRPLFEKKS LEDKTERELLESYIDG	-(R - 0.0034)	4051.9263	4051.9297	0.82	HM
TFGSGEADC(CysGly)GLRPLFEK KSLEDKTERELLESYIDG	+(G + 0.0032)	4265.0489	4265.0520	0.73	HM	
Fibrinogen alpha chain UniProt P02671	<b>K.SSSYSKQFTSSTSYNRGDSTF ESKSYKMADEAGSEADHEGTH STKRGHAKSRP.V</b>	-	<b>5901.70298</b>	<b>5901.70641</b>	<b>0.58</b>	<b>HM</b>
	K.SSSYSKQFTSSTSYNRGDSTF ESKSYKMADEAGSEADHEGTHS TKRGHAKSRP.V	-(V + 0.0010)	5802.63456	5802.6370	0.42	HM
Fibrinogen alpha chain UniProt P02671	<b>K.SSSYSKQFTSSTSYNRGDSTF ESKSYKMADEAGSEADHEGTH STKRGH.A.K</b>	-	<b>5334.3536</b>	<b>5334.3572</b>	<b>0.67</b>	<b>HM</b>
	K.SSSYSKQFTSSTSYNRGDSTF ESKSYKMADEAGSEADHEGTHS TKR.G	-(GHA + 0.0002)	5069.2362	5069.2395	0.65	HM
	K.SSSYSKQFTSSTSYNRGDSTF ESKSYKMADEAGSEADHEGTHS TK.R	-(RGHA + 0.0099)	4913.1351	4913.1288	-1.28	HM

	K.SSSYSKQFTSSTSYNRRGDSTF ESKSYKMADEAGSEADHEGTHS T.K	-(KRGHA - 0.0005)	4785.0401	4785.0442	0.86	HM
Platelet Factor 4 UniProt P02776	F.SAEAEEDGDLQCLCVKTTTSQ VRPRHITSLEVIKAPHCPPTAQLIA TLKNGRKICLDLQAPLYKKIIKKLL ES	+(FASA + 0.0385)	8141.3692	8141.3932	2.94	HM
	F.SAEAEEDGDLQCLCVKTTTSQV RPRHITSLEVIKAPHCPPTAQLIATL KNGRKICLDLQAPLYKKIIKKLLS	+(SA - 0.0027)	7923.2636	7923.2464	-2.17	HM
	S.AEAEEDGDLQCLCVKTTTSQVR PRHITSLEVIKAPHCPPTAQLIATLK NGRKICLDLQAPLYKKIIKKLLS	+(A + 0.0297)	7836.2316	7836.2468	1.93	HM
	<b>A.EAEEDGDLQCLCVKTTTSQVR PRHITSLEVIKAGPHCPPTAQLIAT LKNGRKICLDLQAPLYKKIIKKLL ES</b>	-	<b>7765.1945</b>	<b>7765.1799</b>	<b>-1.88</b>	<b>HM</b>
	E.AEEDGDLQCLCVKTTTSQVRPR HITSLEVIKAPHCPPTAQLIATLKN GRKICLDLQAPLYKKIIKKLLS	-(E + 0.0132)	7636.1518	7636.1240	-3.64	HM
	A.EEDGDLQCLCVKTTTSQVRPRH ITSLEVIKAPHCPPTAQLIATLKN RKICLDLQAPLYKKIIKKLLS	-(EA + 0.068)	7565.1147	7565.0933	-2.83	HM
HMW Kininogen UniProt P01042	<b>H.NLGHGHKHERDQGHGHQ</b>	-	<b>1943.9080</b>	<b>1943.9071</b>	<b>-0.47</b>	<b>LM</b>
	H.NLGHGHKHERDQGHGHQ	+(H + 0.0001)	2080.9669	2080.9661	-0.39	LM
Complement C3f fragment UniProt P01024	<b>SSKITHRIHWESASLLR</b>	-	<b>2021.1039</b>	<b>2021.1040</b>	<b>0.04</b>	<b>LM</b>
	S.SKITHRIHWESASLLR	-(S + 0.0005)	1934.0719	1934.0715	-0.20	LM
	S.KITHRIHWESASLLR	-(SS + 0.0013)	1847.0399	1847.0387	-0.63	LM
	H.RIHWESASLLR	-(SSKITH + 0.0003)	1367.7542	1367.7541	-0.14	LM
	I.HWESASLLR or RIHWESASLL	-(SSKITHRI/L - 0.0005)	1098.5691	1098.5697	0.56	LM
Factor XIIIa UniProt P00488	<b>R.VPPNNSNAEEDDLPTVELQ GVVPR.G</b>	-	<b>2602.3107</b>	<b>2602.3113</b>	<b>0.22</b>	<b>LM</b>
	A.VPPNNSNAEEDDLPTVELQGL VPR.G	-(A - 14.0141)	2545.2893	2545.2883	-0.38	LM
	A.VPPNNSNAEEDDLPTVELQGV VPR.G	-(A - 0.0006)	2531.2736	2531.2735	-0.04	LM

## DISCUSSION

The past decade, MS-based profiling studies have been carried out to determine disease-specific serum peptidome signatures in a “case-control” setting. Due to the relatively high biological variability of the serum peptidome (and proteome) a large number of samples are required for statistical evaluation. Thus, high-throughput analytical methodologies have been adopted in combination with MS, pioneered by SELDI-TOF platforms. In the same period, high-throughput robotic platforms with more flexible and user-defined sample preparation protocols were combined with MALDI-TOF read-out. Both low-resolution TOF-profiles with a wide  $m/z$ -range and high-resolution profiles with smaller  $m/z$ -windows were reported for proteins and peptides, respectively [7;30;31]. However, single- or even multi-step protein fractionations still yield highly complex samples and the low resolving powers in linear mode SELDI- or MALDI-TOF profiles do not allow accurate quantification of the profiled species. Peptides up to  $m/z$ -values of 4500 can be

routinely analysed with isotopic resolution using TOF-analysers in reflectron mode, but at the cost of restricting the analyzed  $m/z$ -range and thus excluding proteins from the evaluation. Moreover, reflectron mode profiles still contain a significant number of overlapping peptides, as we previously demonstrated in ultrahigh resolution MALDI-FTICR profiles [20].

In this study the ultrahigh resolving power provided by a 15 tesla MALDI-FTICR system was exploited in terms of discriminative power of case-control peptidome profiles and identification of observed species. This is the first profiling study that reports on the application of such ultrahigh resolution profiles exemplified by a clinical cohort of serum samples from healthy individuals and PC patients. Aiming for cancer-specific peptide and protein signatures, these serum samples were first fractionated on a fully automated SPE-platform based on functionalized MBs and then profiled using a 15T MALDI-FTICR mass spectrometer. In total, 487 peptides or small proteins (*i.e.* 196 and 291 in LM and HM spectra, respectively) were measured with isotopic resolution in the  $m/z$ -range 1-9 kDa and quantified with high accuracy and precision. The ultrahigh resolving power allowed the correct quantification of peptides or proteins that previously were observed to suffer from overlapping isotopic distributions in lower resolution profiles (see Figure2). Note that the total number of detectable peptides was higher than 487, *i.e.* several peptides were detected only in few particular samples, probably due to a higher expression of a particular protein or an elevated protease activity.

Two different MALDI-FTICR acquisition methods, namely a low mass and a high mass method, were used to generate peptide and protein profiles from two independent groups of serum samples. A calibrated and validated discriminating rule built on the combination of the data obtained from the two MALDI-FTICR methods resulted in a sensitivity of 89% and a specificity of 100% with an AUC of 0.989. These results corroborate classification numbers from our previous MALDI-TOF studies [19;32]. The t-test analysis performed on the peptides with absolute discriminant weights higher than 0.1 resulted in the identification of 34 peptides that (*i.e.* p-value lower than 0.001) differentiate between case and control groups (see Table 2). The high precision and accuracy of the mass measurements allowed the identification of 26 of these peptides either by comparison with previously reported peptides or by accurate mass measurement of mass differences in the spectra (see the Materials and Method section). Application of the latter approach resulted

in the identification of peptides generated through proteolysis of the same protein. In fact, starting from a previously identified peak (*i.e.* peptide) it was found that accurate measurement of the difference between that specific  $m/z$ -value and the  $m/z$ -value of a new peak matched to a similar peptide with either one amino acid more or less at the C- or the N-terminus, corresponding to the “overall” protein sequence. Thus, up to 8 new peptides could be identified starting from the fragment peptide K.SLEDKTERELLESYIDGR of thrombin light chain (UniProt P00734) (see Table 3). Nevertheless, the presence of isobaric peptides cannot be excluded and MS/MS experiments are required to further validate the identifications. In conclusion, using the two identification approaches described above, we are now able to further expand the total number of identified peptides, especially at higher  $m/z$ -values. Other MALDI-profiling methods that so far have been used for the characterization of human serum peptides were not suitable for the identification of high molecular weight peptides or proteins, because these lacked sensitivity and resolving power [28;29].

As a final remark, it should be noted that at this stage the peptidome profiles were not evaluated for the  $m/z$ -range from 9,000 to 10,000. Here, both the high density of peaks and the relatively lower resolving power do not permit binning of the data points. The most abundant peaks present in this range were identified as apolipoprotein-CIII isoforms [26] and these data will be evaluated in a separate study using a different quantification method. In this study, we have shown that high quality human serum peptide and protein profiles can be generated using a standardized and robust protocol for the sample preparation and ultrahigh resolution 15T MALDI-FTICR MS for the mass measurements. The use of this mass analyzer allowed the isotopic resolution and the accurate and precise mass measurement of a high number of peptides and small proteins in a wide  $m/z$ -range. Notably, recent innovation in ICR-cell technology potentially provides similar performance at a lower magnetic field strength [33]. The statistical analysis of profiles generated from a clinical cohort of samples allowed the discrimination between healthy individuals and PC patients with sensitivity and specificity comparable with those reported by other authors using MALDI-TOF MS. A total of 273 serum samples was processed and mass analyzed within a time frame of 24 hours and the high quality of the data both facilitated the interpretation and evaluation of the generated profiles. These ultrahigh resolution mass spectra represent a “next-generation” of

MS-based peptidome profiles and provide a new tool for a more detailed description of the high-abundant proteins in clinical serum sample cohorts aiming for new diagnostic leads.

## SUPPLEMENTARY INFORMATION

**Supplementary Table 1. Peptides present in MALDI-FTICR precision profiles that were used for internal calibration.**

Protein (Uniprot Entry)	Peptide sequence	<i>m/z</i> -value	Method
Fibrinopeptide alpha chain 5-16 (P02671)	G.EGDFLAEGGGVR	1206.5749	LM
Fibrinopeptide alpha chain 3-16 (P02671)	D.SGEGDFLAEGGGVR	1350.6284	LM
Phosphorylated Fibrinopeptide alpha chain (P02671)	ADSpGEGDFLAEGGGVR	1616.6594	LM
Complement C3f fragment (P01024)	SSKITHRIHWESASLLR	2022.1067	LM
Factor XIIIa 14–38 (P00488)	R.AVPPNNSNAEEDLPTVELQGVVPR.G	2603.3136	LM
Fibrinogen alpha chain precursor 576–601 (P02671)	K.SSSYSKQFTSSTS <sup>YN</sup> RGDSTFESKSY.K	2932.2944	LM
Fibrinogen alpha chain 76–604 (P02671)	K.SSSYSKQFTSSTS <sup>YN</sup> RGDSTFESKSYKMA.D	3262.4670	LM
Prothrombin 328-363 (P00734)	TFGSGEADC(C <sub>ys</sub> )GLRPLFEKKSLEDKTERELLESYIDGR	4208.0275	HM
Fibrinogen alpha chain 577–624 ( P02671)	K.SSSYSKQFTSSTS <sup>YN</sup> RGDSTFESKSYKMADEAGSEAD HEGTHSTKRGHA.K	5337.3618	HM
Fibrinogen alpha chain 577–629 ( P02671)	K.SSSYSKQFTSSTS <sup>YN</sup> RGDSTFESKSYKMADEAGSEAD HEGTHSTKRGHAKSRPV.R	5904.7111	HM
Apolipoprotein-CIII ( P02656)	SEAEDASLLSFMQGYMKHATKTAKDALSSVQESQVAQ QARGWVTDGFSSLKDYW STVKDKFSEFWDLDPVVRPTSVAA	8765.2387	HM
Glycosylated Apolipoprotein-CIII ( P02656)	SEAEDASLLSFMQGYMKHATKTAKDALSSVQESQVAQ QARGWVTDGFSSLKDYW STVKDKFSEFWDLDPVVRPT(GalNAcGalneu5Ac)SVAA	9421.4664	HM

## REFERENCES

- [1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: Cancer J. Clin.* 2011;61:69-90.
- [2] Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011. *CA: Cancer J. Clin.* 2011;61:212-236.
- [3] Wong T, Howes N, Threadgold J, Smart HL, Lombard MG, Gilmore I, Sutton R, Greenhalf W, Ellis I, Neoptolemos JP. Molecular Diagnosis of Early Pancreatic Ductal Adenocarcinoma in High-Risk Patients. *Pancreatology* 2001;1:486-509.
- [4] Sener SF, Fremgen A, Menck HR, Winchester DP. Pancreatic cancer: a report of treatment and survival trends for 100,313 patients diagnosed from 1985-1995, using the National Cancer Database. *J. Am. Coll. Surg.* 1999;189:1-7.
- [5] Goonetilleke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur. J. Surg. Oncol.* 2007;33:266-270.
- [6] Nilsson T, Mann M, Aebersold R, Yates JR, Bairoch A, Bergeron JJM. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* 2010;7:681-685.
- [7] Albrethsen J. The first decade of MALDI protein profiling: a lesson in translational biomarker research. *J. Proteomics* 2011;74:765-773.
- [8] de Noo ME, Tollenaar RA, Deelder AM, Bouwman LH. Current status and prospects of clinical proteomics studies on detection of colorectal cancer: hopes and fears. *World J. Gastroenterol.* 2006;12:6594-6601.
- [9] Velstra B, van der Burgt YE, Mertens BJ, Mesker WE, Deelder AM, Tollenaar RA. Improved classification of breast cancer peptide and protein profiles by combining two serum workup procedures. *J. Cancer Res. Clin. Oncol.* 2012;138:1983-1992.
- [10] Callesen AK, Vach W, Jørgensen PE, Cold S, Mogensen O, Kruse TA, Jensen ON, Madsen JS. Reproducibility of Mass Spectrometry Based Protein Profiles for Diagnosis of Breast Cancer across Clinical Studies: A Systematic Review. *J. Proteome Res.* 2008;7:1395-1402.
- [11] Beretta L. Proteomics from the clinical perspective: many hopes and much debate. *Nat. Methods* 2007;4:785-786.
- [12] Anderson NG. Adventures in Clinical Chemistry and Proteomics: A Personal Account. *Clin. Chem.* 2010;56:154-160.
- [13] Kinsinger CR, Apffel J, Baker M, Bian X, Borchers CH, Bradshaw R, Brusniak MY, Chan DW, Deutsch EW, Domon B, Gorman J, Grimm R, Hancock W, Hermjakob H, Horn D, Hunter C, Kolar P, Kraus HJ, Langen H, Linding R, Moritz RL, Omenn GS, Orlando R, Pandey A, Ping P, Rahbar A, Rivers R, Seymour SL, Simpson RJ, Slotta D, Smith RD, Stein SE, Tabb DL, Tagle D, Yates JR, Rodriguez H. Recommendations for mass spectrometry data quality metrics for open access data (corollary to the Amsterdam principles). *Prot. Clin. Appl.* 2011;5:580-589.
- [14] Callesen AK, Christensen Rd, Madsen JS, Vach W, Zapico E, Cold S, Jørgensen PE, Mogensen O, Kruse TA, Jensen ON. Reproducibility of serum protein profiling by systematic assessment using solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2008;22:291-300.

- [15] Bladergroen MR, Derks RJE, Nicolardi S, de Visser B, van Berloo S, van der Burgt YEM, Deelder AM. Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids. *J. Proteomics* 2012;77:144-153.
- [16] Percy A, Chambers A, Yang J, Domanski D, Borchers C. Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 2012;404:1089-1101.
- [17] Nicolardi S, Dalebout H, Bladergroen MR, Mesker WE, Tollenaar RAEM, Deelder AM, van der Burgt, YEM. Identification of Human Serum Peptides in Fourier Transform Ion Cyclotron Resonance Precision Profiles. *Int. J. Proteomics* 2012; 804036.
- [18] Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.* 2006;116:271-284.
- [19] Velstra B, Bonsing BA, Mertens BJ, van der Burgt YEM, Huijbers A, Vasen H, Mesker WE, Deelder AM, Tollenaar RAEM. Detection of pancreatic cancer using serum protein profiling. *HPB* 2013;15:602-610.
- [20] Nicolardi S, Palmblad M, Dalebout H, Bladergroen M, Tollenaar RAEM, Deelder AM, van der Burgt YEM. Quality control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spectrom.* 2010;21:1515-1525.
- [21] Marshall J, Jankowski A, Furesz S, Kireeva I, Barker L, Dombrovsky M, Zhu W, Jacks K, Ingratta L, Bruin J, Kristensen E, Zhang R, Stanton E, Takahashi M, Jackowski G. Human Serum Proteins Pre-separated by Electrophoresis or Chromatography Followed by Tandem Mass Spectrometry. *J. Proteome Res.* 2004;3:364-382.
- [22] Hagman C, Ramström M, Håkansson P, Bergquist J. Quantitative Analysis of Tryptic Protein Mixtures Using Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Proteome Res.* 2004;3:587-594.
- [23] Hertkorn N, Ruecker C, Meringer M, Gugisch R, Frommberger M, Perdue EM, Witt M, Schmitt-Kopplin P. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Anal. Bioanal. Chem.* 2007;389:1311-1327.
- [24] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* 1998;17:1-35.
- [25] Scigelova M, Hornshaw M, Giannakopoulos A, Makarov A. Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* 2011; 10.
- [26] Nicolardi S, van der Burgt YEM, Wührer M, Deelder AM. Mapping O-glycosylation of apolipoprotein C-III in MALDI-FTICR protein profiles. *Proteomics* 2013; 13:992-1001
- [27] Nicolardi S, Palmblad M, Hensbergen PJ, Tollenaar RAEM, Deelder AM, van der Burgt YEM. Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* 2011;25:3457-3463.
- [28] Hortin GL. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.* 2006;52:1223-1237.



[29] Tiss A, Smith C, Menon U, Jacobs I, Timms JF, Cramer R. A well-characterised peak identification list of MALDI MS profile peaks for human blood serum. *Proteomics* 2010;10:3388-3392.

[30] Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P. Serum Peptide Profiling by Magnetic Particle-Assisted, Automated Sample Processing and MALDI-TOF Mass Spectrometry. *Anal. Chem.* 2004;76:1560-1570.

[31] Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P. Correcting Common Errors

in Identifying Cancer-Specific Serum Peptide Signatures. *J. Proteome Res.* 2005;4:1060-1072.

[32] Velstra B, Burgt Y, Mertens B, Mesker W, Deelder AM, Tollenaar R. Improved classification of breast cancer peptide and protein profiles by combining two serum workup procedures. *J. Cancer Res. Clin. Oncol.* 2012;138:1983-1992.

[33] Nikolaev EN, Jertz R, Grigoryev A, Baykut G. Fine Structure in Isotopic Peak Distributions Measured Using a Dynamically Harmonized Fourier Transform Ion Cyclotron Resonance Cell at 7 T. *Anal. Chem.* 2012;84:2275-2283.





# **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 measurement 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

1. Nørregaard Jensen, O. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 2004, 8 (1), 33-41.
2. Buchen, L. Missing the mark. *Nature* 2011, 471 (7339), 428-432.
3. Mitchell, P. Proteomics retrenches. *Nat. Biotech.* 2010, 28 (7), 665-670.
4. Anderson, N. L. Counting the proteins in plasma. *Clin. Chem.* 2010, 56 (11), 1775-1776.
5. Albrethsen, J. The first decade of MALDI protein profiling: a lesson in translational biomarker research. *J. Proteomics* 2011, 74 (6), 765-773.
6. Villanueva, J.; Lawlor, K.; Toledo-Crow, R.; Tempst, P. Automated serum peptide profiling. *Nat. Protocols* 2006, 1 (2), 880-891.
7. Kulasingam, V.; Pavlou, M. P.; Diamandis, E. P. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer* 2010, 10 (5), 371-378.
8. Petricoin III, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 2002, 359 (9306), 572-577.
9. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* 2013, 113 (4), 2343-2394.
10. Patel, B. B.; Barrero, C. A.; Braverman, A.; Kim, P. D.; Jones, K. A.; Chen, D. E.; Bowler, R. P.; Merali, S.; Kelsen, S. G.; Yeung, A. T. Assessment of Two Immunodepletion Methods: Off-Target Effects and Variations in Immunodepletion Efficiency May Confound Plasma Proteomics. *J. Proteome Res.* 2012, 11 (12), 5947-5958.
11. Jones, K. A.; Kim, P. D.; Patel, B. B.; Kelsen, S. G.; Braverman, A.; Swinton, D. J.; Gafken, P. R.; Jones, L. A.; Lane, W. S.; Neveu, J. M.; Leung, H. C.; Shaffer, S. A.; Leszyk, J. D.; Stanley, B. A.; Fox, T. E.; Stanley, A.; Hall, M. J.; Hampel, H.; South, C. D.; de la Chapelle, A.; Burt, R. W.; Jones, D. A.; Kopelovich, L.; Yeung, A. T. Immunodepletion Plasma Proteomics by TripleTOF 5600 and Orbitrap Elite/LTQ-Orbitrap Velos/Q Exactive Mass Spectrometers. *J. Proteome Res.* 2013.
12. Marx, V. Targeted proteomics. *Nat. Meth.* 2013, 10 (1), 19-22.
13. Anderson, L.; Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Mol. Cell. Proteomics* 2006, 5 (4), 573-588.
14. Kuzyk, M. A.; Smith, D.; Yang, J.; Cross, T. J.; Jackson, A. M.; Hardie, D. B.; Anderson, N. L.; Borchers, C. H. Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. *Mol. Cell. Proteomics* 2009, 8 (8), 1860-1877.
15. Razavi, M.; Frick, L. E.; LaMarr, W. A.; Pope, M. E.; Miller, C. A.; Anderson, N. L.; Pearson, T. W. High-Throughput SISCAPA Quantitation of Peptides from Human Plasma Digests by Ultrafast, Liquid Chromatography-Free Mass Spectrometry. *J. Proteome Res.* 2012, 11 (12), 5642-5649.

16. Saldova, R.; Struwe, W. B.; Wynne, K.; Elia, G.; Duffy, M. J.; Rudd, P. M. Exploring the Glycosylation of Serum CA125. *Int. J. Mol. Sci.* 2013, 14 (8), 15636-15654.





SUMMARY  
SAMENVATTING  
ACKNOWLEDGEMENTS  
CURRICULUM VITAE  
PUBLICATION LIST

---





## SUMMARY

The human serum proteome, *i.e.* the complex mixture of all proteins that are expressed in the body and subsequently captured in the blood circulation, is considered as a valuable source of information that can be used to describe the health status of an individual. This complexity, however, poses a challenge to the development of analytical techniques that aim for determination of protein identities and quantities, and which later on could be used routinely in clinical laboratories. In the last years much effort has been made to tackle the sample complexity and several analytical methods have been developed to identify and quantify serum proteins at different concentration levels. Mass spectrometry (MS) has evolved as the method of choice for the analysis of proteins resulting into a new scientific field called MS-based proteomics. In a mass spectrometer molecules are separated and detected in the gas phase after using different technologies for ionization. Over the years, the evolution of mass spectrometers has led to the improvement of both the ionization efficiency, the resolution in the separation of ions, the sensitivity, the precision, the accuracy, and the speed of detection expressed as a mass-to-charge ratio ( $m/z$ ) of positively or negatively charged molecules (*i.e.* ions). Despite these improvements analysis of the human serum proteome is only informative (or possible) when MS is used in combination with sample preparation techniques that decrease serum complexity. The analysis of low abundant proteins is seriously hampered by two facts: (a) the twenty most abundant proteins constitute approximately 95% of the total serum protein content and (b) serum protein concentrations range over ten orders of magnitude. Thus, several analytical approaches have been used to separate and/or concentrate proteins prior to MS analysis. As an alternative of MS-based proteome analysis, serum peptide profiling has been used to find “features” in the enormous mixture of peptides that is generated as a result of proteolysis of the most abundant proteins. It has been found that the extent of the proteolysis can be correlated with certain diseases. Matrix-assisted laser desorption/ionization (MALDI) coupled with time-of-flight (TOF) MS has been the preferred MS platform for the acquisition of such peptide (and protein) profiles because of its high-throughput character. In the early days, unfortunately, these profiles were generated at a low mass resolution resulting in the detection of peptide and proteins as broad peaks over a wide mass range. With the advent of reflectron TOF mass analysers similar profiles could be generated at higher mass resolution allowing the detection of the

isotopic distribution, however at the cost of a narrower mass range. Based on serum peptide and protein profiles, early successes were reported on classification of groups of individuals with a specific disease versus groups of healthy (control) individuals with high specificities and sensitivities,. However, lack of standardization in sample preparation, mass measurement and statistical data processing has so far prevented translation of the discriminant methods into valid clinical assays.

Fourier transform ion cyclotron resonance (FTICR) MS provides (a) ultrahigh mass resolving power that allows the analysis of large proteins and complex mixtures; (b) ultrahigh mass measurement accuracy and precision that allow a more reliable identification of the detected species; (c) a wide dynamic range that is favourable for the detection of low abundant components. Despite these excellent characteristics, MALDI-FTICR MS has been rarely used for peptide and protein profiling of large cohorts of samples. In this thesis, the development and application of novel MALDI-FTICR-MS methods for high-throughput analysis of human serum peptides and proteins is described. Due to the large biological variability of human serum, the analysis of a large cohort is needed to obtain reliable statistical results.

In **Chapter 1** the implementation of various solid-phase extraction (SPE) sample preparation protocols on two different, dedicated platforms is described, namely: 1) Magnetic bead-based SPE of peptides and proteins from body fluids on a Hamilton liquid handling workstation; 2) Cartridge-based SPE on a SPARK Symbiosis system. These two platforms, specifically set-up for high-throughput analysis using optimized protocols, provide fast analytical sample preparation procedures needed for MS-based clinical proteomics. As discussed above, MALDI-TOF has been the most used MS platform for peptide profiling and therefore was used as a reference in our development strategies. To test the performance of a new Bruker solariX<sup>TM</sup> MALDI-FTICR system equipped with a 15 tesla magnet, 96 different serum samples were processed using the magnetic bead-based SPE protocol and measured with ultrahigh resolution up to  $m/z$ -value 6500 by MALDI-FTICR MS and up to  $m/z$ -value 3500 using a Bruker Ultraflex II system. In **Chapter 2**, the results obtained with the two mass spectrometers are described and compared in terms of mass accuracy, mass measurement precision and

repeatability of the signal. The results show that the two systems perform rather similar with respect to the repeatability of the signal intensities, while mass measurement errors (MMEs) are a few ppm's or at a sub-ppm level for TOF and FTICR measurement, respectively. Moreover, the mass measurement precision improves at least 10-fold in ultrahigh resolution data thus simplifying spectral alignment necessary for robust and quantitatively precise comparisons of profiles in large-scale clinical studies. In chapter 2, we furthermore report on additional identifications of reversed-phase (RP) C18-fractionated serum peptides by collision-induced dissociation (CID) experiments performed on the FTICR system.

It is common practice to measure each serum sample on multiple MALDI-spots in order to prevent loss of data in case of unsuccessful MALDI spotting, In **Chapter 3**, we describe the implementation of a new quality control (QC) parameter for both MALDI-TOF and MALDI-FTICR peptide profiling based on isotopic distributions that allows the selection of the best MALDI spectrum from replicate spectra. To test the performance of this QC parameter, RPC18-fractionated human serum peptides were measured by high resolution MALDI-TOF and ultrahigh resolution MALDI-FTICR. For each single mass spectrum, the isotopic distributions of multiple peptides were compared with estimated polyaveragine distributions to calculate a QC value. This value, calculated for replicate spots, allows the selection of the optimal MALDI spectrum and exclusion of low quality spectra for further statistical analysis. Using this approach, spectra containing high intensities of polymers or other contaminants and lacking peptides of interest can be efficiently removed from a clinical dataset.

The benefits of the ultrahigh resolving power and the high mass accuracy and precision provided by an FTICR mass spectrometer equipped with a 15-tesla magnet for the identification of human serum peptides is further shown in **Chapter 4**. Ultrahigh resolution data not only allow the assignment of fragment ions with high charge states (4+, 5+) but also enhance confidence of human serum peptide identifications from tandem MS experiments. This is exemplified with collision-induced dissociation (CID) and electron transfer dissociation (ETD) data of middle-down-sized endogenous or protein-breakdown peptides that are of interest in biomarker discovery studies.

The new MALDI-FTICR methods described in this thesis allow the ultrahigh resolution MS analysis of peptides and small proteins up to  $m/z$ -value 15,300. Apolipoprotein C-III (apoCIII) is a small human serum protein that is expressed as four major isoforms. A detailed evaluation of these isoforms is reported in **Chapter 5**. MALDI-FTICR profiles generated from RPC18-fractionated human serum peptides and proteins are discussed with regard to reproducibility of the signal intensities as well as the accurate mass measurements. Moreover, ESI-FTICR-MS/MS spectra were evaluated to confirm the identity of the apoCIII glycoforms. The reported analytical procedure is well-suited for high-throughput screening of serum samples and provides detailed information of apoCIII isoforms, thus allowing detailed investigation on the role of this protein in various disease pathways.

In **Chapter 6**, we describe the identification of six new apoCIII isoforms characterized by more complex glycan moieties which are fucosylated instead of sialylated. Top-down ESI-FTICR-MS/MS and bottom-up LC-ion trap MS/MS spectra were evaluated to confirm the glycan moiety and localize the glycosylation site. Analysis of these fucosylated isoforms in 96 different serum samples showed a large variation in their presence and abundance. This warrants further research to elucidate the possible implication of these glycoforms in studies where alterations in apoCIII have been linked to the development of different pathologies.

In **Chapter 7** a powerful, fast and robust approach for obtaining serum biomarker signatures based on a fully automated one-step solid-phase extraction serum sample clean-up in combination with fast MALDI acquisition and ultrahigh precision FTICR-readout, is described. The ultrahigh resolution FTICR methods implemented in this analytical procedure allow the measurement with isotopic resolution and low-ppm accuracy of all peptide and proteins between 1 and 9 kDa. Moreover, as exemplified for a cohort of pancreatic cancer patient and controls, the generated ultrahigh resolution MALDI-FTICR spectra can be used to classify serum samples with sensitivity and selectivity values well above 85%.

The last part of this thesis, the **General Discussion**, contains a discussion about MS-based cancer biomarker discovery strategies and the new developments in human serum peptide and protein profiling by ultrahigh resolution MS reported in Chapters 1-7. Targeted proteomics approaches using multiple-reaction monitoring (MRM) MS are briefly discussed in the context of further validation of biomarker candidates. Finally, a growing interest is seen in SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies) MS methods as an alternative for biomarker discovery strategies.

## SAMENVATTING

Het humane proteoom is het complexe mengsel van alle eiwitten die in het lichaam tot expressie worden gebracht. Een groot deel van deze eiwitten komt vervolgens terecht in de bloedsomloop. Dit zogenaamde serumproteoom wordt algemeen beschouwd als een rijke bron van informatie betreffende de gezondheidstoestand van een persoon. De complexiteit van het eiwitmengsel heeft echter tot gevolg dat de ontwikkeling van analytische technieken voor de nauwkeurige identificering en kwantificering van eiwitten voor grote uitdagingen staat, met name wanneer deze methoden later voor routinebepalingen in een klinisch laboratorium ingezet dienen te worden. Massaspectrometrie (MS) heeft zich inmiddels ruimschoots bewezen als meest geschikte technologie voor zulke bepalingen, en “*MS-based proteomics*” is een zelfstandig onderzoeksgebied geworden. In een massaspectrometer worden moleculen in de gasfase gescheiden en gedetecteerd, nadat deze zijn geïoniseerd met behulp van een ionisatiemethode. Moleculen worden gedetecteerd als positief dan wel negatief geladen ionen en het signaal wordt weergegeven in een massaspectrum met een massa-lading ratio ( $m/z$ ) op de x-as en een intensiteit op de y-as. Instrumentparameters die gedurende tientallen jaren van ontwikkeling en vernieuwing een belangrijke rol hebben gespeeld zijn de ionisatie-efficiëntie, het scheidend vermogen (“resolutie”), de gevoeligheid, de nauwkeurigheid (“precisie”), de juistheid (“accuratesse”), en de snelheid van de meting. Voor succesvolle MS-metingen blijft een geschikte monstervoorbewerking cruciaal, dit ondanks de grote stappen die zijn gemaakt op het gebied van instrumentontwikkeling. In deze context is er veel werk verricht om de complexiteit van serummonsters te verkleinen en daarmee eiwitten van uiteenlopende concentraties te kunnen identificeren en kwantificeren. Veel van de duizenden verschillende typen eiwitten in het (serum) proteoom zijn namelijk aanwezig in lage concentraties (milligram, microgram of zelfs nanogram per liter). Bovendien zijn deze eiwitten “verstopt” onder een twintigtal hoog-abundante eiwitten, die samen ongeveer 95% van de totale hoeveelheid eiwitmassa vertegenwoordigen. In plaats van het meten van alle eiwitten in het serumproteoom, een potentieel krachtige maar tijdrovende en kostbare strategie, is het ook mogelijk een enkel peptideprofiel te generen van een serummonster. Veel van deze peptiden zijn het resultaat van een proteolyse van de hoog-abundante eiwitten. In verschillende studies heeft men laten zien dat de omvang van proteolyse, die ten dele weerspiegeld wordt in

peptideprofielen, correleert met specifieke ziekten. Het meest gebruikte MS-platform voor het genereren van peptide- en eiwitprofielen bestaat uit een “*time-of-flight (TOF)*” massa-analysator met een “*matrix-assisted laser desorption/ionization (MALDI)*” ionisatiebron. Aanvankelijk (10-15 jaar geleden) werden peptiden en eiwitten als brede pieken gedetecteerd, omdat de gebruikte apparatuur in deze vroege studies een laag-scheidend vermogen had. De komst van reflectron TOF (10 jaar geleden) betekende een enorme verbetering: hogere resolutie maakte isotoopdistributie van de peptiden zichtbaar, echter ten koste van de gevoeligheid voor eiwitten. Sindsdien hebben verschillende onderzoeksgroepen laten zien dat op basis van serumpeptide- en/of eiwitprofielen patiëntsera met hoge sensitiviteit en specificiteit kunnen worden geclassificeerd in zogenaamde “*case-control*” studies. Echter, door gebrek aan standaardisering van monstervoorbewerking, MS-experimenten en statistische analyses heeft (nog) geen translatie plaatsgevonden van deze discriminerende profielen naar een valide klinische test.

Ultrahoge resolutie “*Fourier transform ion cyclotron resonance (FTICR)*” massaspectrometrie maakt metingen van grote eiwitten in complexe monsters mogelijk door het ultrahoog scheidend vermogen. Daarnaast zorgen de verbeterde precisie en accuratesse ervoor dat gedetecteerde peptiden en eiwitten met grotere zekerheid kunnen worden geïdentificeerd. Tenslotte maakt het grote dynamisch bereik van een FTICR-analyse deze techniek zeer geschikt voor het detecteren van componenten met een lage concentratie. Desalniettemin is de combinatie van MALDI en FTICR nog slechts zelden ingezet voor de analyse van een cohort. In dit proefschrift beschrijven wij de ontwikkeling en de toepassing van nieuwe MALDI-FTICR-MS methoden voor de “*high-throughput*” (hoge-doorvoer) analyse van humane serumpeptiden en –eiwitten. Het is belangrijk te vermelden dat serummonsters veel verschillen tussen verschillende individuen vertonenn (biologische variatie), zodat analyse van grote cohorten essentieel is voor het verkrijgen van betrouwbare statistische data. De implementatie van verschillende protocollen voor monstervoorbewerking gebaseerd op “*solid-phase*” (vaste-fase) extractie (SPE) is beschreven in **Hoofdstuk 1**. Twee platforms zijn ingezet: 1) SPE met behulp van gefunctionaliseerde paramagnetische bolletjes op een Hamilton pipetteerrobot en 2) SPE met behulp van cartridges op een SPARK Symbiosis systeem. Beide platforms, specifiek



opgezet voor “*high-throughput*” analyses met geoptimaliseerde protocollen zorgen voor een snelle monstervoorbewerking, noodzakelijk voor “*MS-based clinical proteomics*”. Voor het verkrijgen van peptideprofielen werd een Bruker Ultraflex II MALDI-TOF gebruikt als referentie (tot  $m/z$  3500) en daarnaast een Bruker solariX™ MALDI-FTICR systeem, uitgerust met een 15 tesla magneet (tot  $m/z$  6500). In **Hoofdstuk 2** zijn de verkregen profielen met elkaar vergeleken voor wat betreft accuratesse, precisie van de massabepaling en herhaalbaarheid van de meting met betrekking tot intensiteiten. MALDI-TOF en MALDI-FTICR waren op dit laatste punt gelijkwaardig. Echter, de andere twee parameters waren meer dan 10x beter op het ultrahoge resolutie systeem, waarmee het op één-lijn-brengen van alle profielen in een cohort eenvoudiger en robuuster wordt. Daarnaast zijn in **Hoofdstuk 2** nieuwe identificaties van serum peptiden gerapporteerd.

In de meeste studies worden van ieder serummonster meerdere MALDI-spots (“replica’s”) gemaakt ten einde kwaliteitscontrole mogelijk te maken. In **Hoofdstuk 3** is een nieuw type kwaliteitscontrole beschreven voor peptideprofielen. Deze parameter is gebaseerd op isotoopdistributies, waarmee voor elke meting bepaald kan worden wat het beste profiel van een monster is binnen een set replica’s. Profielen, of spectra, van slechte kwaliteit kunnen op deze manier worden uitgesloten van verdere statistische analyses. Daarna is in **Hoofdstuk 4** opnieuw aangetoond dat het gebruik van ultrahoge resolutie FTICR-MS zeer geschikt is voor identificatie van (nieuwe) serumpeptiden. Hiervoor zijn twee fragmentatietechnieken gebruikt, namelijk “*collision-induced dissociation (CID)*” en “*electron transfer dissociation (ETD)*”. Vervolgens zijn de MALDI-FTICR methoden ingezet voor het verkrijgen van peptide- en eiwitprofielen van humane serummonsters tot een  $m/z$ -waarde van 15300. In de profielen zijn verschillende isovormen van het eiwit apolipoproteïn C-III aanwezig, zoals beschreven in **Hoofdstuk 5** van dit proefschrift. Het “*profielen*” van dit eiwit, met daarbij de isovormen met hun verschillende glycaanstructuren, is veelbelovend voor het monitoren van een bepaalde ziekte waarbij apolipoproteïnen een belangrijke rol spelen. In **Hoofdstuk 6** is verder onderzoek beschreven aan apolipoproteïn C-III. In dit hoofdstuk zijn zes nieuwe isovormen gerapporteerd, waarbij de glycanen fucose-eenheden bevatten in plaats van siaalzuren. Deze structuren zijn bevestigd met behulp van additionele “*bottom-up LC-ion trap*”

*MS/MS-experiments*". De analyse van deze nieuwe structuren in 96 verschillende serummonsters gaf een gedetailleerd inzicht in de biologische variatie die gepaard gaat met glycosylering van eiwitten in het algemeen en fucosylering op apolipoproteïn C-III in het bijzonder. Tenslotte is in **Hoofdstuk 7** een studie beschreven, waarbij de ontwikkelde, krachtige "*high-throughput*" MALDI-FTICR-MS strategie in combinatie met één-staps-monstervoorbewerking (gefunctionaliseerde paramagnetische bolletjes) ingezet is voor de analyse van een cohort van serummonsters van patiënten met alvleesklierkanker. Aan de hand van peptideprofielen met isotoopresolutie tot 9 kDa konden patiënten van controles worden onderscheiden met sensitiviteits- en specificiteitswaarden hoger dan 85%, voor zowel een calibratie- als een validatieset. In het laatste deel van dit proefschrift, de **Algemene Discussie**, wordt kort vooruitgeblikt op "*multiple-reaction monitoring (MRM) MS*" methoden, die kunnen dienen voor het verder valideren van de biomarker kandidaten die in dit proefschrift beschreven staan.

## ACKNOWLEDGEMENTS

The journey of my life continues now toward new horizons and new ambitions. I will always look back to the years of my PhD with emotion, aware of how much the experiences and feelings of this period have shaped what I am now. During this time, I have met many different people who contributed directly or indirectly to this thesis. Here, I want to thank them all because, thanks to them, my PhD has been a very pleasant and interesting experience.

I would like to thank Yuri, my co-promotor. Your contribution to this thesis is immense. You helped, taught and supported me. You have been a great mentor and a good friend. As I wrote in my propositions at the beginning of this thesis, opportunities are the key for personal development and success. For this reason, I will always be grateful to André, my promotor, who gave me the opportunity to join his group to do research with the best available technology. You are a visionary scientist and with your successful career and interpersonal skills inspired my work and my life.

If I had to process all my samples manually, probably my PhD would have lasted forever. Thanks Marco. Your robotic arms speeded up my research.

Statistics can be complicated when you are not an expert. Thanks Bart for your help and for the time you spent explaining me the kind of analysis that you had just performed.

During my PhD I often needed some help to solve small issues. I always encountered a colleague who was willing to 'waste' part of his/her time for me. So, thanks to all of them for their support and help.

Thanks Bart, Katja, Kate and Paul for the extra fun that we had together after work.

Kate, you are the perfect office mate. Thanks for your happiness that made me smile every day. Thanks also to all the other people that shared the same office with me. I really enjoyed the time we spent together.

Paul, thank you very much. Your great passion for science and your scientific curiosity inspired me and triggered my imagination for new experiments and studies.

Thanks Hans for your help and for teaching me the basic words of Dutch.

The first year of my PhD was tough, then I met Antonio, David, Mireia, Ceci, Raul and Gloria. With your Spanish vitality you made me happy. Thank you very much chicos.

Rafael and Lara thank you very much for the time we spent together. We lived in the same house as brothers and sisters, we had fun and we shared much of ourselves including cutlery and chairs. You will always be in my heart.

La comunità italiana a Leiden è davvero dinamica. Qui ho incontrato tante persone provenienti da diverse regioni Italiane. Grazie Erica, Alessio, Manu, Antonietta ed Elena per aver speziato la mia italianità con la vostra cultura.

Ciccio, Irene, Pietro, Simona e Valentina, insomma a Famiglia. Siete entrati così velocemente nella mia vita da non farmi capire quando ho iniziato a volervi così tanto bene. Grazie per tutto quello che avete fatto per me, Tiziana e Diego.

Un grazie particolare lo rivolgo a mia madre, a mio padre, a miei fratelli Gabriele e Giovanni ed a mia sorella Elisa. Anche se siete lontani vi sento tutti molto vicino e questo mi dà sicurezza. Il successo del mio dottorato di ricerca è anche merito vostro.

Vivere quest'avventura con una persona speciale al fianco è stato fantastico. Abbiamo condiviso veramente tutto ed alla fine le gioie e dolori ci hanno unito ancora di più. Grazie Tiziana per il tuo supporto. Questa tesi è anche tua.

Diego, grazie per la tua gioia e voglia di vivere. I tuoi sorrisi hanno illuminato le giornate negli ultimi anni del mio dottorato.

## **CURRICULUM VITAE**

Simone Nicolardi was born on June 28, 1981, in Gagliano del Capo (LE), Italy. He studied chemistry at the technical secondary school “ITIS Pacinotti” of Taranto and at the faculty of Chemistry at the University of Rome “Sapienza”. Here, under the supervision of Prof. Dr. A. Di Corcia, he graduated in 2006 with a thesis on the development of liquid chromatography-tandem mass spectrometry (LC-MS) for monitoring quinolone residues in bovine milk. After his graduation, he worked for almost one year, as analytical chemist, at the Italian Institute of the Environmental Protection and Research (ISPRA) where he analyzed persistent organic pollutants in sea sediments and fishes using gas and liquid chromatography coupled with mass spectrometry and other detectors. In February 2008, he moved to Almeria, Spain. Here, he attended the laboratory of Prof. Dr. A. R. Fernandez-Alba working on the development of LC-MS and ELISA methods for the analysis of pharmaceutical residues in both wastewater and surface water. In December 2008, he started his PhD studies at the Department of Parasitology at the Leiden University Medical Center (LUMC), the Netherlands, under the supervision of Prof. Dr. A.M. Deelder, Prof. Dr. R.A.E.M. Tollenaar and Dr. Y.E.M. van der Burgt. The PhD research resulted in this thesis entitled “Development of ultrahigh resolution FTICR mass spectrometry methods for clinical proteomics”. Since December 2012 he works as post-doc in the Center for Proteomics and Metabolomics at the LUMC continuing his research on the development of ultrahigh resolution MS-based methods for the analysis of peptides and proteins.

## LIST OF PUBLICATIONS

1. **Nicolardi S**, Deelder AM, Palmblad M, and van der Burgt YEM. On-line electrochemical reduction of disulfide bonds facilitates structure analysis of an intact monoclonal antibody using ultrahigh resolution mass spectrometry. *Submitted for publication.*
2. **Nicolardi S**, Velstra B, Mertens BJ, Bonsing B, Mesker WE, Tollenaar RAEM, Deelder AM, and van der Burgt YEM. Ultrahigh resolution profiles lead to more detailed serum peptidome signatures of pancreatic cancer. *Translational Proteomics* 2014, 2, 39-51.
3. **Nicolardi S**, Giera M, Kooijman P, Kraj A, Chervet J, Deelder AM, and van der Burgt YEM. On-line electrochemical reduction of disulfide bonds: improved FTICR-CID and -ETD coverage of oxytocin and hepcidin. *J. Am. Soc. Mass Spect.* 2013, 24, 1980-1987.
4. Jónasdóttir HS, **Nicolardi S**, Jonker W, Derks R, Palmblad M, Ioan-Facsinay A, Toes R, van der Burgt YE, Deelder AM, Mayboroda OA, Giera M. Detection and structural elucidation of esterified oxylipids in human synovial fluid by electrospray ionization-Fourier transform ion-cyclotron mass spectrometry and liquid chromatography-ion trap-MS(3): detection of esterified hydroxylated docosapentaenoic acid containing phospholipids. *Anal. Chem.* 2013, 85, 6003-6010.
5. **Nicolardi S**, van der Burgt YEM, Dragan I, Hensbergen PJ, and Deelder AM. Identification of new apolipoprotein-CIII glycoforms with ultrahigh resolution MALDI-FTICR mass spectrometry of human sera. *J. Proteome Res.* 2013, 12, 2260-2268.
6. **Nicolardi S**, van der Burgt YE, Wuhler M, Deelder AM. Mapping O-glycosylation of apolipoprotein C-III in MALDI-FT-ICR protein profiles. *Proteomics* 2013, 13, 992-1001.
7. Gianazza E, Chinello C, Mainini V, Cazzaniga M, Squeo V, Albo G, Signorini S, Di Pierro SS, Ferrero S, **Nicolardi S**, van der Burgt YE, Deelder AM, Magni F. Alterations of the serum peptidome in renal cell carcinoma discriminating benign and malignant kidney tumors. *J. Proteomics* 2012, 76, 125-140.

8. Bladergroen MR, Derks RJE, **Nicolardi S**, de Visser B, van Berloo S, van der Burgt YEM, and Deelder AM. Standardized and automated solid-phase extraction procedures for high-throughput proteomics of body fluids. *J. Proteomics* 2012, 77, 144-153.
9. **Nicolardi S**, Herrera S, Martínez Bueno MJ and Fernández-Alba AR. Two new competitive ELISA methods for the determination of caffeine and cotinine in wastewater and river waters. *Anal. Methods* 2013, 4, 3364-3371.
10. **Nicolardi S**, Dalebout H, Bladergroen MR, WE Mesker WE, Tollenaar RAEM, Deelder AM, and van der Burgt YEM. Identification of human serum peptides in Fourier transform ion cyclotron resonance precision profiles. *Int. J. Proteomics* 2012, 804036.
11. **Nicolardi S**, Andreoni A, Tabares LC, van der Burgt YEM, Canters GW, Deelder AM and Hensbergen PJ. Top-down FTICR MS for the identification of fluorescent labeling efficiency and specificity of the Cu-protein Azurin. *Anal. Chem.* 2012, 84, 2512-2520.
12. **Nicolardi S**, Palmblad M, Hensbergen PJ, Tollenaar RAEM, Deelder AM and van der Burgt YEM. Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Comm. Mass Spectrom.* 2011, 25, 3457-3463.
13. **Nicolardi S**, Palmblad M, Dalebout H, Bladergroen M, Tollenaar RAEM, Deelder AM, van der Burgt YEM. Quality control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling. *J. Am. Soc. Mass Spect.* 2011, 21, 1515-1525.
14. Bogialli S, D'Ascenzo G, Di Corcia A, Laganà A, **Nicolardi S**. A simple and rapid assay based on hot water extraction and liquid chromatography-tandem mass spectrometry for monitoring quinolone residues in bovine milk. *Food Chem.* 2008, 108 (1), 354-360.