Automated analysis of mouse serum peptidome using restricted access media and nanoliquid chromatography–tandem mass spectrometry

Geun-Cheol Gil, Jim Brennan, Dan J. Throckmorton, Steven S. Branda, Gabriela S. Chirica*

Systems Biology Department, Sandia National Laboratories, Livermore, CA 94551, USA

ABSTRACT

We demonstrate use of restricted access media with reversed phase functionality (RAM-RP) for analysis of low molecular weight proteins and peptides in mouse serum (75 μl) using a custom designed modular automated processing system (MAPS). RAM-RP fractionation with simultaneous removal of high molecular weight and high abundance proteins is integrated with a follow-on buffer exchange module (BE) to ensure compatibility with subsequent processing steps (trypsin digestion and intact peptide separation prior to mass spectrometric analysis). The high sample capacity afforded by chromatographic methods generates enough sample to achieve comprehensive serum peptidome identification (357 proteins) through tandem mass spectrometric analysis of both intact and digested peptides. Sample losses during transfer between modules are minimized through precise fluidic control; no clogging occurred over several months of serum processing in our low back pressure system. Computer controlled operation of both modules and thorough optimization yield excellent run-to-run reproducibility and protein/peptide overlap in analytical repeats. The robustness of our results demonstrate that the RAM-RP-BE workflow executed on our MAPS platform shows tremendous potential for high throughput peptidome processing, particularly with regard to direct analysis of small-volume serum samples.

Published by Elsevier B.V.

1. Introduction

In mammalian bodily fluids, the low molecular weight (LMW) proteome, otherwise known as the peptidome, is comprised of functional small proteins and peptides such as hormones, cytokines, chemokines and growth factors [1–4], as well as proteolytic products of high molecular weight (HMW) proteins from serum and body organs. It has been suggested that processes such as crosspresentation (replacement of exogenous peptides bound to major histocompatibility complex molecules by endogenous peptides) and peptide sequestration by blood carrier proteins can generate and, respectively, preserve and concentrate peptides [5,6]. Recent studies utilizing mass spectrometry for structural identification or quantitative profiling of the peptidomes revealed that bodily fluid samples constitute vast sources of endogenous peptides [1–7]. Information about the peptide or parent protein structure, tissue origin, alterations in concentration, posttranslational modifications, and specific proteolytic cleavage patterns can support biomarker studies for disease diagnosis, treatment, and monitoring [6–8]. There is, therefore, great potential for discovery of peptidome profiles that support discrimination between normal and disease states.

Analytical challenges associated with processing of complex bodily fluid samples are currently limiting the clinical application of peptidome analysis for biomarker discovery. To begin with, the peptidome represents only a small fraction of the total amount of proteins in bodily fluids. For instance, about 95% of the total serum proteome consists of high abundance proteins which are, for the most part, high molecular weight (HMW) proteins. They are present in up to ten orders of magnitude higher concentrations, overwhelming the signal of lower abundance LMW proteins [9]. Effective removal of these high abundance proteins typically depends upon immunodepletion methods that are costly and dilute the sample concentration up to twenty fold. Another challenge stems from the large number and wide structural variety of peptides and small proteins in bodily fluids, such that extensive sample fractionation and concentration must be carried out in order to meet the mass and dynamic range capabilities of mass spectrometers, avoid ionization suppression and undersampling to ultimately improve peptidome coverage [3,5]. This typically entails laborious processing and many sample transfer, freezing and thawing steps, prolonged exposure to proteases resulting in sample loses and poor reproducibility. Such limitations are particularly damaging when only small sample volumes are available, as in small animal studies and time-course analyses. There is a clear need for high throughput selective fractionation methods that enable reliable and comprehensive profiling of bodily fluid peptidomes.

* Corresponding author. Tel.: +1 925 294 4876; fax: +1 925 294 3282.
E-mail address: gschiri@sandia.gov (G.S. Chirica).

Several approaches to selective peptidome sampling from complex bodily fluids have been demonstrated. Merrel et al. used organic acids to precipitate and remove large proteins, enabling dissociation of small proteins from the larger abundance carrier molecules to reveal less abundant species [10]. Ultrafiltration with 10–30 kDa filters has been used frequently for selective removal of high abundance, larger proteins [11–14] though some leakage of high molecular weight molecules and great variance in cut-off accuracy has been noted [15]. Electrophoretic methods [16] and size exclusion chromatography [17] also have shown promise in this regard but, similar to ultrafiltration, the small sample load and low throughput have limited their utility in biomarker discovery applications.

Readily automated and reproducible, chromatographic methods are ideal for high-throughput processing. Either in continuous or solid-phase extraction format they reduce sample complexity and improve identification of lower abundance components. The high loading capacities of chromatographic processing significantly improve concentration factors. Lower volume, concentrated fractions can be readily processed in scalable, miniaturized formats to take advantage of limited diffusion, and increased sensitivity. Furthermore, there is a wide range of existing chromatographic methods and materials which enable concentration of specific classes of proteins or peptides and, thereby, divide complex samples into manageable fractions. Operating these methods in on-line, automatable format makes chromatographic workflows a good match for high throughput peptidome research.

Restricted access media (RAM) is typically used for concentration of small molecules in complex biological samples and recently it has found new application in the analysis of the peptidomes in bodily fluids [18–21]. RAM is a chromatographic sorbent based on a porous silica or organic polymer. The outer surface of the packing is hydrophilic and non-adsorptive. The pores have small diameters, and are functionalized with ionic or hydrophobic groups. This provides a medium which combines size exclusion chromatography with partition or ion-exchange chromatography: low-mass molecules are retained in the pores, whereas larger molecules are excluded and do not stably interact with the inert outer surface, instead eluting in the void volume [19,21]. Wagner et al. designed a 2-D system for mapping the peptidomes of human hemofiltrates and cell lysates; the peptides are concentrated on RAM with ion exchange functionality, then transferred and further deconvoluted on four reversed phase columns, yielding fractions ready for off-line MALDI-TOF MS analysis [22]. Further improvement in throughput is achieved in the system built by Hu et al. [23]; the eluent of a RAM cation exchange (RAM-CatEx) cartridge is loaded on a capillary liquid chromatography (LC) column directly coupled to nanoLC–tandem mass spectrometry. RAM-CatEx materials have also been synthesized in monolithic format for direct integration of sample clean-up [24]. Column life time for analysis urine and plasma samples is significantly improved [22–24]. The dual chromatographic capabilities of RAM materials make them ideal for direct injection of complex samples for automated analysis of the peptidomes.

In this study we demonstrate use of a modular automated processing system (MAPS) [25–27] for continuous analysis of small-volume (75 μl) mouse serum samples. In our system, RAM-RP and buffer exchange (BE) modules integrate selective separation of the serum peptidome and delivery of fractions in matrices/buffers ready for digestion and mass spectrometric analysis. Customized and miniaturized valves interconnected to cartridges and low-pressure pumps confer the precise fluidic transfer between modules and the low dead volumes required for processing microliter samples. Automation, real-time monitoring and rapid prototyping enable thorough system optimization required for highly reproducible serum analysis. Seamless integration of the RAM-RP-BE modules demonstrates that MAPS can significantly expand the exploratory and clinical applications of peptidome research.

2. Experimental

2.1. Materials and reagents

Bovine serum albumin (BSA), ammonium bicarbonate, iodoacetamide (IAA), formic acid, and dithiothreitol (DTT) were purchased from Sigma–Aldrich (St. Louis, MO). Acetonitrile, methanol, and HPLC grade water were obtained from EMD Chemicals (Gibbstown, NJ). Trypsin (sequencing grade) was from Promega (Madison, WI). Pooled CD 1 mouse serum was purchased from Innovative Research (Novi, MI). BioGelP2 size exclusion packing was purchased from BioRad (Hercules, CA). Bond Elut Plexa, (Varian, Palo Alto, CA) is the RAM utilized in this study; we refer to it as RAM-RP, to highlight the reverse-phase separation mechanism and distinguish it from other RAM materials with different functionalities. Seppro IgY-M7 is an immunodepletion material designed to remove the 7 most abundant proteins from mouse serum (albumin, IgG, IgM, transferrin, haptoglobin, fibrinogen and α1-antitrypsin) and was purchased from GenwayBio (San Diego, CA).

2.2. MAPS-enabled fractionation of mouse serum proteins

MAPS [25,27] joins commercial and customized hardware components (valves, pumps, cartridges, UV detector, autosampler and fraction collector) to perform on-line processing and fractionation of complex biological samples. The Spark autosampler (Spark, Emmen, The Netherlands), the NE-500 OEM syringe pumps (New Era Pump Systems, Wantagh, NY), the UVVIS 200 Linear detector (Reno, NV), the Rheodyne selection valve (IDEX, Oak Harbor, WA) and the BioRad 100 fraction collector (BioRad, Hercules, CA) are commercially available. Custom-designed components include miniaturized electronically actuated 3-way valves [28], PEEK cartridges of 2 μl to 2.5 ml internal volumes [25], CapTite™ or TubTite™ fittings for direct capillary interconnection of all system components, and a breadboard platform for temporary or permanent hardware positioning in complex configurations [29]. The PEEK cartridges can be readily refilled with any type of packing, conferring a wide functionality range to the MAPS modules [25]. Custom software enables rapid communication with individual modules for an essentially unlimited number of serial and parallel processing steps. The MAPS platform bridges micro and mesofluidic applications (10–10,000 μl) and has been used for sample processing in fully integrated systems for environmental analysis [25–27].

The MAPS workflow built for the analysis of mouse serum proteins is depicted in Fig. 1. Six 3-way valves connected to the RAM-RP and BE columns were mounted on the prototyping breadboard. The small footprint of the valves and cartridges, and the close capillary connections between them, minimizes dead-volumes and reduces band broadening. Two Rheodyne valve toggle between ports as New Era syringe pumps refill with select buffers or infuse the buffers onto cartridges during conditioning, washing, running and cleaning steps. The first processing module fractionates serum using a 200 μl cartridge packed with RAM-RP. Following a conditioning step, the autosampler injects the sample (serum diluted 1:3 with 0.1% formic acid in water) directly on the RAM-RP module at 50 μl/min flow rate. The cartridge is washed, and elution solvents (0.1% formic acid containing various amounts of methanol) are loaded in the syringe pump then flushed through the cartridge at 100 μl/min flow rate as specified in the software. The timing of fraction collection is achieved by coordinating the peak elution (monitored continuously on the UV detector), with the BioRad vial
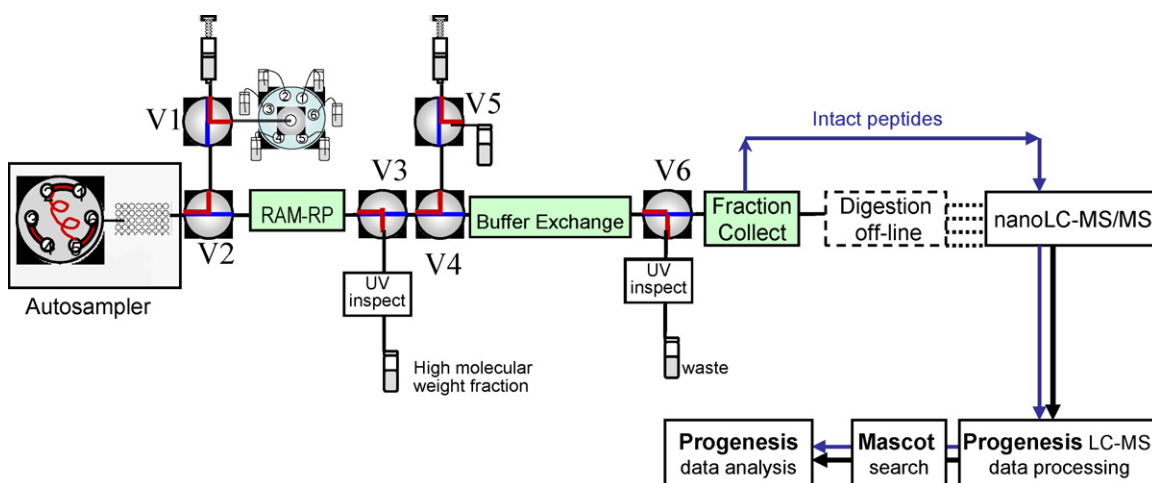
Fig. 1. Schematic diagram of the MAPS setup incorporating the RAM-RP and buffer exchange modules. On-line automated steps are represented with a full line, while off-line processing steps (e.g. trypsin digestion) are represented with dotted lines.

switch. When the BE module is used, valves V3 and V4 are switched as fraction plugs elute off the cartridge, and the eluates are transferred to the 1 ml cartridge packed with BioGel P2. The exchange buffer composition is 50 mM ammonium bicarbonate (pH 8) with 5% methanol. In between runs, the two modules are cleaned and conditioned to minimize carry over and improve reproducibility. The collected fractions are analyzed in the nanoLC–MS/MS setup, with or without prior trypsin digestion.

2.3. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Fractions collected from RAM-RP columns were analyzed using SDS-PAGE. Protein concentrations in the starting serum and RAM-RP fractions were measured using the Bradford assay (BioRad, Hercules, CA). 5 μg of each sample were loaded onto a NuPAGE® Novex 4–20% Tris–Glycine gel (Invitrogen, Carlsbad, CA) and after electrophoresis the gel was stained with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA) or, for improved the detection sensitivity, with SilverExpress (Invitrogen, Carlsbad, CA).

2.4. Trypsin digestion of proteins

Protein fractions generated in the MAPS setup were reduced, alkylated, and digested with trypsin to enable further analysis in the nanoLC–MS/MS system. Fractions eluted from the RAM-RP cartridge were buffer exchanged in 50 mM ammonium bicarbonate (pH 8) with 5% methanol, using the size exclusion cartridge (Fig. 1). The proteins were reduced by mixing with DTT (5 mM final concentration) and incubating the sample at 60 °C for 1 h. After cooling on ice, IAA was added to a final concentration of 10 mM, and the sample was incubated at room temperature in the dark for 30 min. The reduced and alkylated proteins were then digested with trypsin at 37 °C overnight. The reaction was quenched with 1% (v/v) formic acid, yielding peptides in 0.1% (v/v) formic acid ready for nanoLC–MS/MS analysis.

2.5. NanoLC-ESI-MS/MS analysis

Separation and identification of the peptides and proteins present in the processed samples was achieved using a nanoLC system coupled to an ESI Q-TOF Ultima mass spectrometer (Waters, Milford, MA). Peptide separation was achieved on an Eksigent nanoLC 2D system (Eksigent, Dublin, CA) equipped with a trap column onto which the sample was loaded at 1 μl/min and a

separation column was operated at 300 nl/min. The trap column (ID = 250 μm, L = 3 cm) was packed with Macherey Nagel C18, particles of 7 μm diameter. The loaded peptides were gradually eluted and resolved onto a pulled tip capillary (ID = 75 μm, L = 15 cm) packed with Zorbax C18 (Agilent, Santa Clara, CA) of 3.5 μm diameter. The mobile phases were 0.1% (v/v) formic acid in 2/98% (v/v) acetonitrile/water (solvent A) and 0.1% (v/v) formic acid in 98/2% v/v acetonitrile/water (solvent B). Peptide separation on the pulled capillary was achieved as solvent B was gradually increased from 5% to 30% over 100 min, then to 70% over 45 min. The column was re-equilibrated for 20 min prior to each run.

We used [Glu]-Fibrinopeptide B human peptide (Sigma, St. Louis, MI) in order to evaluate the detection limits on our LC–MS system. According to our observation, the limit of detection on the LC–MS was calculated 0.2 fmol/μl of the peptide with signal to noise ratio 3:1. We spiked [Glu]-Fibrinopeptide in our sample and were able to see the peptide with the same signal to noise. The limit of quantification was measured 1 fmol/μl with less than 10% CV.

The eluent was continuously electrosprayed into the Q-TOF instrument equipped with an orthogonal Z-spray. The source temperature was 80 °C. The capillary voltage was 2.8 kV and the cone voltage was 100 V. MS spectra were acquired across the mass range of m/z 350–1900. Automatic function switching from MS to MS/MS was performed when the intensity of individual ion rose above an intensity threshold of 25. Survey scans were acquired during 0.9 s with an inter-scan delay of 0.1 s. MS/MS to MS switchback criteria were defined as followed: TIC rising above threshold of 3000 counts/s or after 6 s. MS/MS scans time was 1.9 s and inter-scan delay was 0.1 s. Three product ion scans were collected for each cycle and parent ions were excluded from further selection for 60 s.

2.6. MS data analysis

The raw data files generated using the Q-TOF mass spectrometer were loaded to Progenesis LC–MS version 3.1 (Nonlinear Dynamics, UK) and the MS and MSMS spectra were transformed to peak list. One replica was selected as a reference and the retention times of all other replicas were aligned by automatic alignment to maximize the 2D feature overlay. Features with only one or more than five charges were excluded to minimize false positive peptide match. The LC–MS/MS replicas were grouped into the corresponding MAPS-processed sample sets, and the raw abundance of all features was normalized. One-way analysis of variance (ANOVA) was calculated over all features in all samples using normalized

abundance. The MS/MS peak lists were converted as a Mascot Generic File (mgf) by Progenesis LC–MS and then loaded onto Mascot server version 3.3 (Matrix Science, UK). Searching parameters were setup to search the SwissProt 2010 database (selected for Mus musculus, 16246 entries) assuming the digestion enzyme as trypsin for tryptic digested peptides or as non-specific for endogenous peptides. Mascot was searched with a fragment ion mass tolerance of 0.5 Da and a parent ion tolerance of 300 ppm. One missed cleavage was allowed in Mascot search for tryptic peptide search. Iodoacetamide derivative of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine or phosphorylation of serine and threonine were specified in Mascot as variable modifications. False positive ratio was $\leq 2.5\%$ for tryptic peptides and $\leq 4.2\%$ for intact peptides by a Mascot-integrated decoy database search with an ion score cut-off of 30 and a significance threshold of $p \leq 0.01$. The resulting proteins containing at least one unique peptide with ion score of 30 and above was exported as an xml file from Mascot and imported back to Progenesis LC–MS for the identification of proteins and peptides. The peptides and proteins list were exported to Excel to evaluate the reproducibility of sample sets and LC–MS repeat runs.

The protein and peptide match list from Mascot was also exported as DAT file and then imported to Scaffold version 3.1 (Proteome Software Inc., Portland, OR) in order to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95% probability as specified by the Peptide Prophet algorithm [30]. Protein identifications were accepted if they could be established at greater than 90% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm [31]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone, were grouped to satisfy the principles of parsimony.

3. Results and discussion

3.1. Optimization of RAM-RP module

Restricted access media (RAM) materials are ideal for high-throughput peptidome analysis. We used Bond Elut Plexa, a RAM with reverse-phase functionality (RAM-RP), for enrichment of small proteins and peptides from mouse serum samples. The size-restrictive hydrophobic core of this material is sampled only by smaller analytes, such as peptide and small proteins, which are later eluted with buffers containing organic solvents. The outer surface of the material is hydrophilic to prevent binding of larger proteins, which elute in the void volume. An advantage of the Bond Elut Plexa packing for serum analysis lies in the hydrophobic retention which allows direct injection of samples with high salt concentration. The material is designed for solid-phase extraction application, with particle sizes ranging from 20 to 40 μm; the 10 μm membranes used in our cartridges retain the material while allowing unrestrictive flow of samples. We operated the RAM-RP material at low pressures (100 psi) to protect the polymeric packing and enabled use of inexpensive low-pressure pumps. As a result, the operation of RAM-RP in our platform was extremely robust: we ran over a hundred samples with no sign of clogging.

A 200 μl RAM-RP cartridge was incorporated in the MAPS platform as shown in Fig. 1. Pooled CD1 mouse serum (75 μL) was diluted 1:3 with 0.1% formic acid to reduce sample viscosity and thereby facilitate access to the hydrophobic pores of the RAM. The sample was loaded directly on the conditioned RAM-RP module and, after wash steps, small proteins and peptides were eluted with increasing concentration of methanol (10–80%) in 0.1% formic acid. These washing and cleaning steps combined with specific hardware
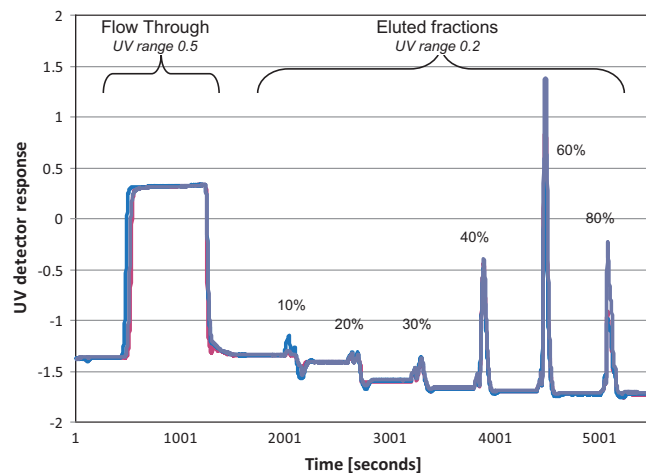

Fig. 2. Overlay of 3 chromatograms in which 500 μl sample (serum diluted 1:3 in 0.1% formic acid) was fractionated on a 200 μl volume cartridge packed with RAM-RP (Bond Elut) material. Elution buffers were 10%, 20%, 30%, 40%, 60%, and 80% methanol in 0.1% formic acid.

design [25,28,29] offered robust, clog-free operation for over 120 runs per cartridge during four months of testing. The excellent reproducibility afforded by automated operation of conditioning, washing and cleaning steps is reflected in the overlay of three chromatograms shown in Fig. 2. Another advantage of automation and real-time monitoring (accomplished here with a UV detector), is the ability to adjust the elution buffer compositions on the fly to generate predetermined amounts of proteins per fraction and/or a more uniform peak area distribution within the fractions as required in specific applications. In our initial experiments (Fig. 2) we expected peptides and small proteins to elute in the low methanol fractions and aimed to matched the eluted protein amount with the sample mass requirement for nanoLC–MS/MS analysis (about 20–30 μg for 3 repeats). However, the SDS PAGE of all the fractions indicated that a substantial amount of small proteins are also present in the fractions eluted with higher methanol concentrations (Fig. 3). In subsequent experiments, we considered the constraints of our high-throughput application, and improved the efficiency in use of MS instrument time by rationally reducing the number of fractions to be analyzed and adjusted the elution protocol to yield 3 fractions of approximately similar amounts of protein (Fig. 5). In both experiments, reproducible elution time (under 6% relative standard deviation for 9 runs) allowed tight synchronization between peak elution and collection; this improved sample recovery and minimized dilution.

The molecular size distribution of the eluted proteins is a measure of RAM selectivity for the small proteins. Fig. 3 illustrates SDS-PAGE of the whole mouse serum (lane 3) and the four fractions eluted with 30, 40, 60, and 80% methanol (in 0.1% formic acid in water) from the RAM-RP module (lanes 5, 6, 7, and 8). Each lane was loaded with 5 μg of sample, corresponding to about 0.12 μl starting volume of whole serum for lane 3 and 2.5 μl starting serum sample for lanes 5, 6, 7, and 8. The distribution of the protein bands in Fig. 3 shows that proteins and peptides of less than 30 kDa were considerably enriched in all four fractions eluted from the RAM-RP module. A significant proportion of the peptides were quite hydrophobic, given that they eluted with the 60–80% methanol buffers. A number of higher molecular weight proteins co-eluting in the RAM-RP fractions are shown in the silver stained gel. This could be due to the presence a small number of hydrophobic pores with larger inner diameters; it is also possible that the outer hydrophilic surface of the RAM-RP is less inert than anticipated. However, the concentrations of these larger proteins are substantially smaller than those
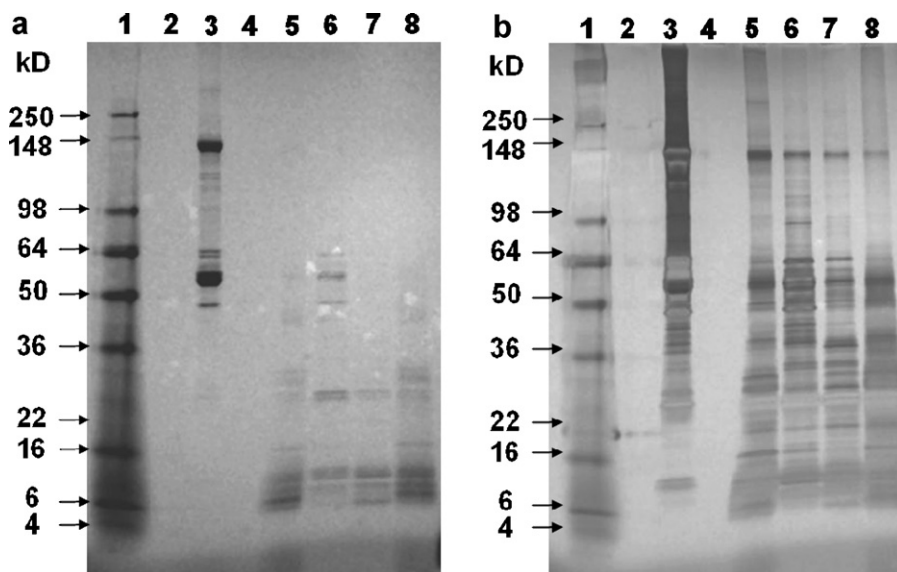
Fig. 3. SDS PAGE of RAM-RP fractions and whole serum with (a) Coomassie and (b) silver staining. Lane 1, marker; lane 2, blank; lane 3, mouse serum; lane 4, blank; lane 5, RAM-RP 30% methanol fraction; lane 6, RAM-RP 40% methanol fraction; lane 7, RAM-RP 60% methanol fraction; lane 8, 80% RAM-RP methanol fraction.

in the whole serum. To further compare the HAP removal efficiency of RAM-RP with immunodepletion methods we packed a cartridge with Seppro IgY-M7 and integrated it in the MAPS setup. Following manufacturer's instructions, we loaded serum diluted 20 fold on the Seppro IgY-M7 cartridge and collected the flow through. Fig. 4 illustrates the SDS PAGE gel loaded with 5 μg of each of the following samples: lane M molecular weight marker, lane 1 whole serum (5 μg originate from approximately 0.1 μl serum), lane 2 immunodepleted serum (5 μg flow-through originate from 0.4 μl serum)

and pooled RAM-RP fractions (5 μg originate from 2.5 μl serum). If we consider that the protein loaded in lane 3 comes from six times the amount of serum used to generate the immunodepleted sample loaded in lane 2, we can conclude that the HAP removal efficiency of the two methods is comparable. Further proof of successful reduction of HAP is the low intensity and small number of albumin and IgG peptides obtained by mass spectrometric analysis (Supplementary Tables 1 and 2). Provided that these larger molecules do not interfere significantly with peptidome detection and identification, the RAM-RP acts as a normalization filter that achieves effective concentration of the smaller proteins, also providing information on proteins in the 30–65 kD range.

The suitability of RAM-RP material for peptidomic applications also depends on its sample capacity and robustness for repetitive injections of unprocessed biofluids such as serum. To prevent sample losses it is important to ensure that the maximum capacity of the inner pores is not exceeded. We assessed the capacity of the Bond Elut RAM-RP material for serum analysis by loading various volumes of samples (diluted 1:3 with 0.1% formic acid in water) on a 200 μl cartridge. The eluted peaks were recorded using the UV detector and the relative peak areas were estimated using custom peak analysis software [26]. Increasing the volume injected on the RAM-RP cartridge from 100 to 300 μl led to non-linear increases in peak areas; based on these results, we determined that the optimal sample volume to inject on a 200 μl cartridge is 200 μl (about 75 μl serum). In these conditions, the relative standard deviations of peak areas from run-to-run and cartridge-to-cartridge were less than 7% and 12%, respectively. The sample capacity is smaller than typical solid phase extraction sorbents because retention relies solely on the inner pores, rather than interaction with all surfaces of the packing material. However, RAM-RP accomplishes two separation modes simultaneously (size exclusion and hydrophobic interaction) such that comparison with immunodepletion materials is more appropriate. Based on our experience, a 200 μl of RAM-RP module can process 75 μl of serum (diluted to 200 μl before loading on the module), whereas a similar sized immunodepletion cartridge typically processes only 10 μl serum (diluted to 300 μl before loading on the module). On the whole, RAM-RP is the material of choice for peptidome analysis not only because of its sample capacity and HAP depletion efficiency, but also its long lifetime, and robust and reproducible performance and long lifetime.
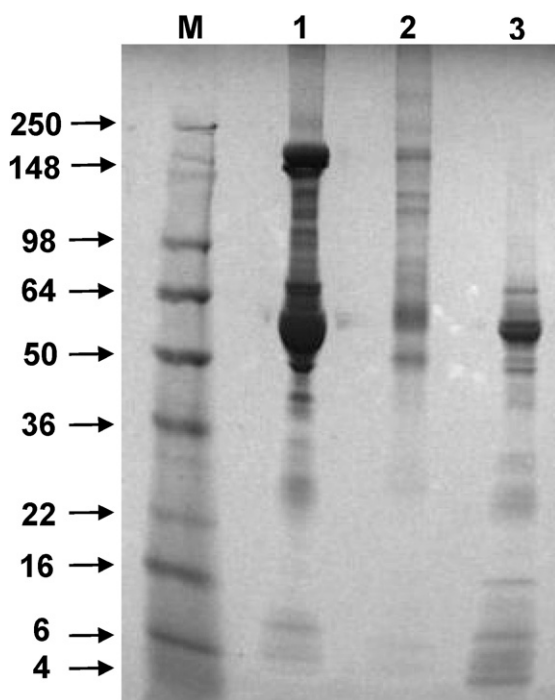

Fig. 4. SDS PAGE gel stained with Coomassie blue illustrates the relative protein size distribution in the samples (5 μg protein mass each) loaded as follows: lane M contains the molecular weight marker, lane 1 contains 0.1 μl whole serum, lane 2 contains flow-through of 0.4 μl serum passed through immunodepletion Seppro IgY-M7 cartridge and lane 3 contains RAM-RP pooled fractions generated from 2.5 μl serum.

3.2. Integration of the RAM-RP and BE modules

Automation of multidimensional fractionation is challenging due to the incompatibility between the specific buffers conditions required for each separation technique. Few separation methods can be directly coupled together. The combination of reversed phase with ion exchange separation is the most common example, and it is often used as the final processing step preceding mass spectrometric analysis. Most processing steps needed to prepare proteins and peptides from crude sample are not directly compatible. To built in-line, automatable workflows the sample passing from one step to the other must be brought in the appropriate format/buffer. For example, RAM-RP and other reversed phase sorbents retain hydrophobic proteins in the context of aqueous solvents and release them in the presence of organic solvents. On the other hand, enzymes typically used in protein digestion are active in very narrow matrix conditions (specific pH range, low salt, very low organic solvent content). Consequently, the RAM-RP fractions cannot be digested immediately following elution; they are typically desiccated under vacuum and re-solubilized, or buffer exchanged using spin columns and centrifugation, in order to meet the conditions for optimal trypsin activity. Thus, the ability to adjust buffer conditions in-line is critical to ensure compatibility between separation conditions and, thereby, enable automation of multidimensional processing.

In our MAPS platform we incorporated a buffer exchange module which receives fractions from the RAM-RP module and changes out the buffer for one (50 mM ammonium bicarbonate, 5% methanol in water, pH 8) that is compatible with tryptic digestion and nanoLC–MS/MS analysis. We packed a 1.7 ml cartridge with BioGel P2 and mounted it in the setup as shown in Fig. 1. We ran bovine serum albumin and tryptophan standards to identify the elution volumes on the column. The BE module can process injection volumes of up to 500 µl. Since the proteins are excluded from the pores and elute in the void volume, their dilution is minimal. Operating parameters for integration of RAM-RP elution with the injection on BE were established following independent optimization of each module. The timing of the RAM-RP peaks elution (switched valve V3 in Fig. 1) was precisely matched with their injection (switched valve V4) on the buffer exchange columns. The timing of the elution of buffer exchanged peaks (switched valve V6) was then matched with the fraction collector. The UV signal of the eluted peaks was recorded following RAM-RP fractionation (Fig. 5). The overlay of the three chromatograms shown in Fig. 5 reflects not only the excellent reproducibility after two processing dimensions, but also the fact that the sample loss during buffer exchange is minimal (approximately 5% reduction in peak area); the observed loss is mostly due to the fact that valve V3 was switched before the complete elution of the small tail of the RAM-RP peaks, to minimize the overall dilution of the collected peak. A significant result is the small overlap between proteins identified in each fraction as shown in Fig. 6. Only fourteen proteins were common to two fractions and thirteen proteins were overlapped in three or more fractions. This suggests the reduced carry over in our automated setup. The system was operated over several months without clogging, despite processing of salty samples as well as solvents with high organic content. This is due to a processing protocol which includes conditioning, washing and cleaning steps of modules and syringes with gradual changes in solvent compositions. Equally important is the replacement of porous metal or PEEK packing retainers, typically used in other chromatographic columns/cartridges, with nylon membranes with 10 µm openings. The robust, reproducible operation of the two dimensional separations demonstrate once more the advantage of automation and the suitability of the MAPS platform for multiprocessing.
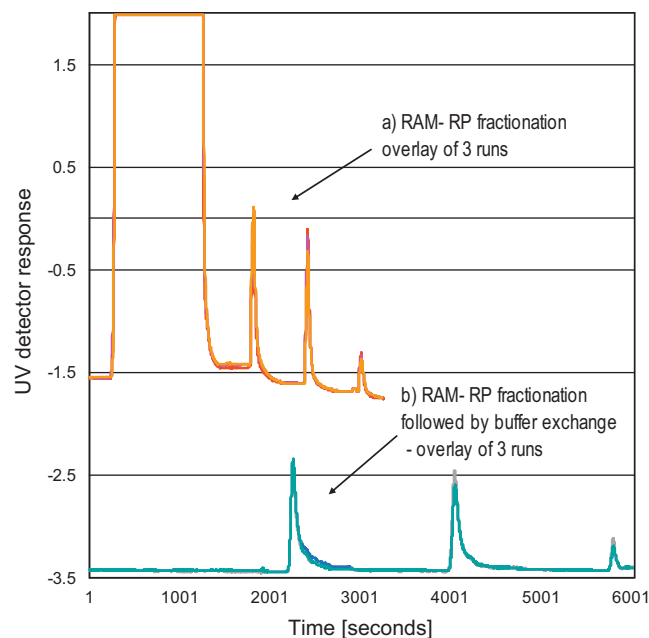

Fig. 5. Overlay of chromatograms recorded following separation on (a) the RAM-RP module and (b) the RAM-RP-BE modules. RAM-RP fractions were eluted using 50, 70 and 90% methanol in 0.1% formic acid.

Structural identification of the peptides and small proteins selected during the RAM-RP-BE fractionation was achieved by nanoLC–MS/MS analysis. RAM-RP-BE fractionation in our platform generates enough protein mass to analyze both digested and intact peptides using nanoLC–MS/MS and thereby extend peptidomic coverage. The MS/MS peak list was generated using Progenesis LC–MS and searched on Mascot as described in the Section 2. To avoid "wonder" hits, the presence of a given protein is typically ascertained if two peptides unique to that protein are identified. This rule introduces a bias toward larger molecular weight proteins for which the probability of identifying more peptides is higher. To compensate for this bias and provide a complete description of the peptidome (which includes small proteins and endogenous peptides) we consider all protein matches, including single peptide hits, and reduce the probability of "wonder" hits by running three LC–MS/MS repeats for each fraction and counting only the peptides identified in two out of three repeats. The supplemental information, provided in Supplementary Tables 1 and 2, lists all proteins and corresponding peptides identified through analysis of digested and intact peptides, respectively. From a total of 357 proteins detected, 132 were identified from 496 tryptic peptides (each detected in two out of three repeats) and 228 were identified from 232 intact peptides (each detected in two out of three repeats). Interestingly, only three proteins were detected common to both methods. This might
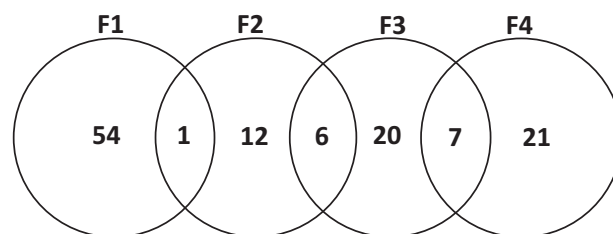

Fig. 6. Distribution of identified proteins among RAM-RP fractions F1, F2, F3 and F4, generated by elution buffers containing 30, 40, 60 and 80% methanol in 0.1% formic acid, respectively. Thirteen proteins were identified from more than three fractions and not shown here to avoid any confusion in the figure.

Table 1
Distribution of the biological functions of the identified proteins following RAM-RP-BE processing from tryptic and intact peptides.

Class of proteins	Proteins identified from tryptic peptides	Proteins identified from intact peptides
Transport, receptor, binding proteins	27	65
Enzymes	25	34
Nuclear proteins, transcription factors	25	38
Structural membrane proteins	18	23
Regulatory proteins	12	23
Growth factors, hormones	6	7
Circulating proteins	6	0
Proteases	4	4
Uncharacterized	3	34
Coagulation, complement factors	3	0
Protease inhibitors	3	0

be due to the fact that small (<2.5 kDa) endogenous peptides are overwhelmed by the large amount of peptides generated following trypsin digestion. The number of identified proteins and peptides is comparable to other reports of serum peptidome analysis: Zheng et al. identified 111 proteins from 361 intact peptides [32] and Hu et al. detected 351 proteins from intact peptide analysis [23].

Reproducibility and reliability of the sample processing and mass spectrometric analysis is essential in all proteomic applications. For a given sample, the most significant sources of variability are the sample processing steps and the separation-through-peak picking steps during LC–MS/MS analysis. Sample processing is mostly manual, time consuming and limited by sample scarcity. Consequently, the latter variability source is typically addressed and minimized by running analytical LC–MS/MS repeats [33,34]. Nonetheless, a minimal number of repeats are desired due to cost and throughput considerations. To identify the number of LC–MS/MS repeats which provide reliable coverage of proteins and peptides in our system, we performed six repeat LC–MS/MS analyses of the tryptic peptides generated in one processing run. Fig. 7(a)

Fig. 7. (a) Trend of total protein identification versus number of iterations. The total number of identified proteins is reported for each repeat of the LC–MS analysis. White bars indicate proteins identified in the previous repeats while dark bars represent the new identifications. Identified proteins (b) from tryptic peptides and (c) intact peptides from three sample sets processed on MAPS with RAM-RP-BE.

depicts the total number of proteins identified in each repeat. The new proteins identified in repeat $n + 1$ against repeat n is colored in black. In the first repeat we identified 115 of proteins (89%), eleven of new proteins (8%) in the second, and two new proteins (3%) in the third repeat. No new protein was identified in the fourth, fifth and sixth repeats. Accordingly, three LC–MS/MS repeats provide comprehensive identification of the proteins present in the tryptic digested fractions. These results are remarkably better than those reported by Camerini et al. which ran a total of eight repeats: three repeats identified about 70% of the proteins, whereas seven repeats were required to identified 98% of the proteins [16]. In the follow-on experiments, we ran three LC–MS/MS repeats for each fraction of each sample/experiment set.

Through automated manipulation of small sample volumes, MAPS enables assessment and, ultimately, reduction of the variability introduced by the sample processing steps. In three experimental/technical replicates the same pooled mouse serum sample was processed through the RAM-RP-BE workflow. For each fraction of each technical replicate we performed three LC–MS/MS repeats, a total of 36 (3 experiments × 4 fractions × 3 repeats) LC–MS/MS runs. Fig. 7(b) and (c) illustrates the excellent overlap between three sample sets for the proteins identified from tryptic and intact peptides, respectively. 76% of the total proteins from tryptic peptides and 47% of the total proteins from intact peptides were identified in all of the three replicate experiments. This is a significant improvement from another serum peptidome study which used ultrafiltration to enrich the peptidome and identified only 15% of the total proteins in three technical replicates [33]. The significant sample-to-sample reproducibility reflects the robustness of RAM-RP workflow as well as the benefit of automation for eliminating variations introduced by manual operation.

For MS data processing and analysis, we employed Progenesis LC–MS and Mascot search engine as described in the Section 2. All ANOVA p values of the identified proteins were less than 0.1 (data not shown), indicating that sample processing on MAPS and peptide analysis on nanoLC–MS/MS are quite reliable. Fig. 8 shows Progenesis LC–MS 3-D montages of the identified tryptic peptide (QSENVGLSSELNR) matched with Testis-specific serine/threonine-protein kinase (accession # TSSK1_Mouse) from three repeat runs of three sample sets. The peak intensity of each repeat from the sample sets are similar, indicating that the sample processing using MAPS and the analysis using LC–MS/MS are quite reproducible.

The proteins identified from tryptic and intact peptides cover a wide spectrum of biological functions as shown in Table 1. Ninety-two (25%) of the identified proteins are transporter, receptors and binding proteins. Sixty-three (17%) are nuclear proteins including transcription factors and fifty-nine proteins (16%) are enzymes including kinases, and transferases. From the biomarker relevance standpoint, we identified a number of peptides which originate from proteins reported to be candidate disease markers: centrosomal protein (retinal degeneration) [35]; histidine ammonia-lyase (histidinemia) [36]; DNA-dependent pro-
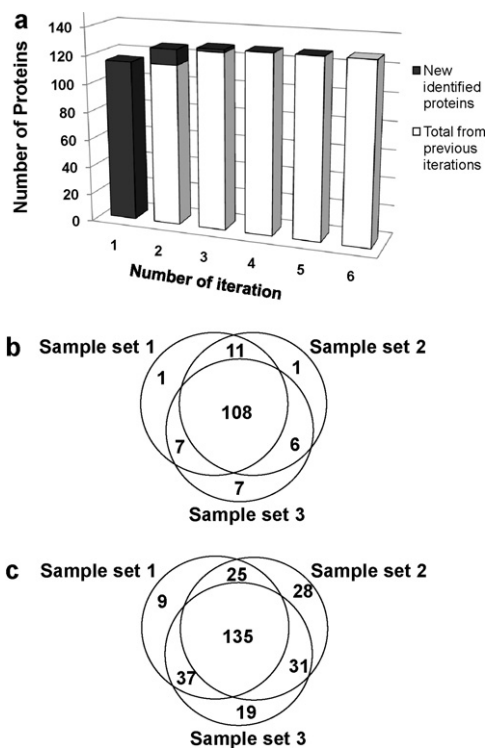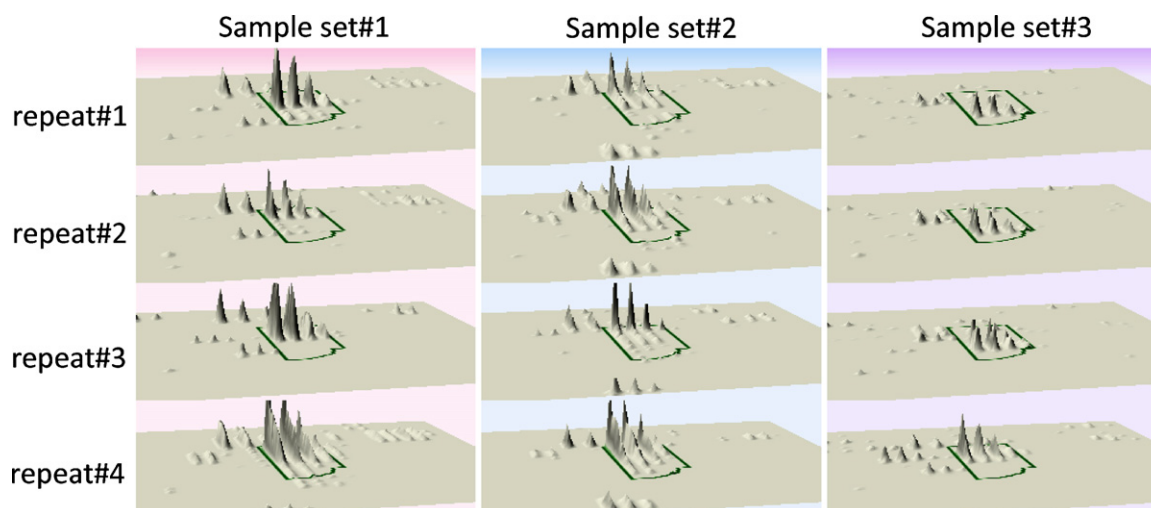
Fig. 8. 3-D montages of the peptide QSENVGLSSELNR of Testis-specific serine/threonine-protein kinase (accession # TSSK1_Mouse) on Progenesis LC–MS. The peptide was generated by trypsin digestion and ANOVA test was performed for the tree sample sets. The p value was 0.06 for the peptide.

tein kinase catalytic subunit (severe combined immune deficiency) [37]; and phosphatidylinositol transfer protein alpha isoform (the vibrator phenotype)[38]. We also identified apolipoprotein A-I and serum amyloid A-3 which were reported as potential biomarkers for diagnosis of acute-phase inflammation [39,40]; counterparts are present in human serum [16,23,32]. These candidate biomarkers were detected all three MAPS experimental replicate, and, using our MAPS platform, can be surveyed in future discovery studies of diseased or infected mouse serum.

4. Conclusions

Mouse serum processing using restricted access media is a rapid approach for peptidome analysis for clinical applications. Its potential for biomarker discovery studies depends not only on throughput but also, most importantly, on the reproducibility and reliability of the analysis. Automated operation confers standardized processing, reduces contamination, enables analysis of small volumes, and ensures operator biosafety during analysis of samples from infectious disease studies. MAPS is a robust platform which not only offers these advantages, but also can bridge incompatible processing steps. Now that we demonstrated the reproducibility and performance of RAM-RP-BE we can add more processing steps, such as on-line digestion and phosphopeptide enrichment, to achieve fully automated processing of serum. Platforms which enable comprehensive, high throughput and time course analysis of bodily fluids in a cost-effective manner can establish proteomic mass spectrometric survey as the method of choice for personalized diagnostics, monitoring and treatment.

Acknowledgement

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2011.03.028.

References

[1] G.V. Baggerman, P. Cleynen, E. Huybrechts, J. De Loof, A.L. Schoofs, J. Chromatogr. B 803 (2004) 3.
[2] M. Fälth, K. Sköld, M. Norrman, M. Svensson, D. Fenyö, P.E. Andren, Mol. Cell. Proteom. 5 (2006) 998.
[3] M. Schrader, P. Schulz-Knappe, Trends Biotechnol. 19 (2001) S55.
[4] M.S. Svensson, K. Svenningsson, P.P.E. Andren, J. Proteome Res. 2 (2003) 213.
[5] O. Chertov, J.T. Simpson, A. Biragyn, T.P. Conrads, T.D. Veenstra, R.J. Fisher, Expert Rev. Proteom. 2 (2005) 139.
[6] L.D. Fricker, J. Lim, H. Pan, F.-Y. Che, Mass Spectrom. Rev. 25 (2006) 327.
[7] L.A. Liotta, E.F. Petricoin, J. Clin. Invest. 116 (2006) 26.
[8] F.L. Strand, Prog. Drug Res. 61 (2003) 1.
[9] N.L. Anderson, N.G. Anderson, Mol. Cell. Proteom. 1 (2002) 845.
[10] K. Merrell, K. Southwick, S.W. Graves, M.S. Esplin, N.S. Lewis, C.D. Thulin, J. Biomol. Technol. 15 (2004) 238.
[11] J. Villanueva, A.J. Martorella, K. Lawlor, J. Philip, M. Fleisher, R.J. Robbins, P. Tempst, Mol. Cell. Proteom. 5 (2006) 1840.
[12] J. Villanueva, D.R. Shaffer, J. Philip, C.A. Chaparro, H. Erdjument-Bromage, A.B. Olshen, M. Fleisher, H. Lilja, E. Brogi, J. Boyd, M. Sanchez-Carbayo, E.C. Holland, C. Cordon-Cardo, H.I. Scher, P. Tempst, J. Clin. Invest. 116 (2006) 271.
[13] S. Mian, S. Ugurel, E. Parkinson, I. Schlenzka, I. Dryden, L. Lancashire, G. Ball, C. Creaser, R. Rees, D. Schadendorf, J. Clin. Oncol. 23 (2005) 5088.
[14] R.J. Robbins, J. Villanueva, P. Tempst, J. Clin. Oncol. 23 (2005) 4835.
[15] F.S. Berven, A.C. Kroksveen, M. Berle, T. Rajalahti, K. Flikka, R. Arneberg, K.M. Myhr, C. Vedeler, O.M. Kvalheim, R.J. Ulrik, Proteom. Clin. Appl. 1 (2007) 699.
[16] S. Camerini, M.L. Polci, L.A. Liotta, E.F. Petricoin, W. Zhou, Proteom. Clin. Appl. 1 (2007) 176.
[17] L.P. Aristoteli, M.P. Molloy, M.S. Baker, J. Proteome Res. 6 (2007) 571.
[18] W.R.G. Baeyens, G. Van der Weken, E. Smet, A.M. García-Campaña, J.P. Remon, J. Pharm. Biomed. Anal. 32 (2003) 913.
[19] N.M. Cassiano, V.V. Lima, R.V. Oliveira, A.C.d. Pietro, Q.B. Cass, Anal. Bioanal. Chem. 384 (2006) 1462.
[20] K. Kishida, N. Furusawa, J. Chromatogr. A 1028 (2004) 175.
[21] A.B. Rudolphi, K.S. Boos, LC-GC 15 (1997) 814.
[22] K. Wagner, T. Miliotis, G. Marko-Varga, R. Bischoff, K.K. Unger, Anal. Chem. 74 (2002) 809.
[23] L. Hu, K.-S. Boos, M. Ye, R.a. Wu, H. Zou, J Chromatogr. A 1216 (2009) 5377.
[24] E. Machtejevas, S. Andracht, D. Lubda, K.K. Unger, J. Chromatogr. A 1144 (2007) 97.
[25] G. Chirica, J. Lachmann, J. Chan, Anal. Chem. 78 (2006) 5362.
[26] J.C. Stachowiak, E.E. Shugard, B.P. Mosier, R.F. Renzi, P.F. Caton, S.M. Ferko, J.L. Van der Vreugde, D.D. Yee, B.L. Haroldsen, V.A. Vandernoot, Anal. Chem. 79 (2007) 5763.
[27] V.A. Vandernoot, R.F. Renzi, B.P. Mosier, J.L. Van de Vreugde, I. Shokair, B.L. Haroldsen, Electrophoresis 31 (2010) 2632.
[28] R.F. Renzi, Microvalve, patent number 6,918,573, application number 10/351,714, 2005.
[29] R.F. Renzi, Microfluidics prototyping platform and components, US 2006/0171852 A1, application number 11/049,378, 2006.
[30] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Anal. Chem. 74 (2002) 5383.
[31] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, Anal. Chem. 75 (2003) 4646.
[32] X. Zheng, H. Baker, W.S. Hancock, J. Chromatogr. A 1120 (2006) 173.
[33] H. Liu, R.G. Sadygov, J.R.Y. III, Anal. Chem. 76 (2004) 4193.
[34] Y. Shen, J.M. Jacobs, D.G.C. II, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, Anal. Chem. 76 (2004) 1134.

[35] B. Chang, H. Khanna, N. Hawes, D. Jimeno, S. He, C. Lillo, S.K. Parapuram, H. Cheng, A. Scott, R.E. Hurd, J.A. Sayer, E.A. Otto, M. Attanasio, J.F. O'Toole, G. Jin, C. Shou, F. Hildebrandt, D.S. Williams, J.R. Heckenlively, A. Swaroop, Hum. Mol. Genet. 15 (2006) 1847.
[36] A.F. Wright, G. Bulfield, S.M. Arfin, H. Kacser, Biochem. Genet. 20 (1982) 245.
[37] R. Araki, A. Fujimori, K. Hamatani, K. Mita, T. Saito, M. Mori, R. Fukumura, M. Morimyo, M. Muto, M. Itoh, K. Tatsumi, M. Abe, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 2438.
[38] B.A. Hamilton, D.J. Smith, K.L. Mueller, A.W. Kerrebrock, R.T. Bronson, V. van Berkel, M.J. Daly, L. Kruglyak, M.P. Reeve, J.L. Nemhauser, T.L. Hawkins, E.M. Rubin, E.S. Lander, Neuron 18 (1997) 711.
[39] Y. Ren, J. Wang, J. Xia, C. Jiang, K. Zhao, R. Li, N. Xu, Y. Xu, S. Liu, J. Proteome Res. 6 (2007) 2812.
[40] C. Martinez, Y. Juarranz, C. Abad, A. Arranz, B.G. Miguel, F. Rosignoli, J. Leceta, R.P. Gomariz, J. Leukoc. Biol. 77 (2005) 729.