



Comparison of serum fractionation methods by data independent label-free proteomics



D. Baiwir^a, G. Mazzucchelli^{b,1}, N. Smargiasso^{b,1}, F. Quesada-Calvo^c, E. De Pauw^b, M. Malaise^d, E. Louis^c, M.-A. Meuwis^{c,*}

^a Proteomic Facility, GIGA, University of Liège, Belgium

^b Laboratory of Mass Spectrometry, GIGA-R, Chemistry Department, CART, ULg, Belgium

^c Gastroenterology Department, GIGA-R, Liège University Hospital, CHU-ULg, Belgium

^d Rheumatology Department, GIGA-R Liège, University Hospital, CHU-ULg, Belgium

ARTICLE INFO

Article history:

Received 26 January 2015

Received in revised form 1 July 2015

Accepted 3 July 2015

Available online 9 July 2015

Keywords:

Serum prefractionation

Repeatability

Label-free proteomics

Data independent

Depletion

Equalization

ABSTRACT

Off-line sample prefractionations applied prior to biomarker discovery proteomics are options to enable more protein identifications and detect low-abundance proteins. This work compared five commercial methods efficiency to raw serum analysis using label-free proteomics. The variability of the protein quantities determined for each process was similar to the unprefractionated serum. A 49% increase in protein identifications and 12.2% of reliable quantification were obtained. A 61 times lower limit of protein quantitation was reached compared to protein concentrations observed in raw serum. The concentrations of detected proteins were confronted to estimated reference values.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The research of proteomic clinical biomarker was often undertaken on serum or plasma to highlight biomarkers available at the systemic level. Indeed clinical trials often collect serum samples which are sometimes available for ancillary biomarkers discovery studies. The low-concentration protein range of serum contains proteins as secretion or tissue leakage products. These are expected to be more specific and sensitive potential disease biomarkers than more abundant proteins [1–3]. The main issue inherent to the complexity of serum analysis is the limitations of the technologies used for the discovery analysis: limits of detection (LOD) and quantification (LOQ) and the linear range of response. Indeed, serum and plasma are very complex biological matrices in which the protein concentration covers a large dynamic range higher than 10 orders of magnitude [4]. Currently no high end mass spectrometer can cover such a large range [5]. Prefractionation strategies offer an alternative by modifying the original sample protein distribution inducing shifting/shrinking effects and allowing proteins initially present at low concentrations to be

accessible for analysis. Several commercial solutions for sample prefractionation are proposed to remove abundant proteins which saturate signal during proteomic analysis and also for reaching the low-abundance proteins [2,6–8]. Of course, off-line or on-line fractionation steps enhance the number of proteins identified by increasing the separation power. But this is at the cost of longer acquisition times, a modified limit of quantification and a higher global variability; all due to the additional preprocessing steps applied [9]. Such multistep strategy is certainly better adapted for cell lines or animal models complete proteome characterization without downstream differential analysis [10]. However for differential analysis and therefore for related clinical proteomics, a higher number of consistent protein identifications and more accurate quantification increase the probability to highlight a significant potential biomarker while keeping reasonable data acquisition time and run length. Therefore, sample preprocessing steps appear mandatory, but must involve at least a good repeatability, reasonable cost and manageable processing time for allowing a medium to high number of clinical sample preparations.

Among the strategies available are the immuno-affinity depletion of the abundant proteins as simple IgG and/or albumin or the depletion of up to the twenty most abundant plasma proteins using IgY chicken antibodies. These IgY decrease the risk for aspecific protein co-depletions potentially occurring with mammalian

* Corresponding author.

E-mail address: marie-alice.meuw@ulg.ac.be (M. -A. Meuwis).

¹ Both the authors contributed equally to this work.

antibodies based depletion [11–15]. Another option is based on the principle of the equalization technique, utilizing a random synthetic hexapeptides library cross-linked to micro-beads [16–21] which actually appears to work according to a general hydrophobic binding mechanism [4]. The large volume of sample which is required and the fact that the “equalization mechanism” shrinks the sample protein concentration dynamic range, can be seen as a problem for data interpretation after differential proteomic discovery analysis and further validation of the results. The enrichment of glycoproteins on lectines may also be used as an alternative approach. Many studies were performed on various systems dedicated to differential analysis [12,22–24]. However, beside works done for protocols testing [14,20,23,25,26], only one study comparing plasma preprocessing reproducibility using two commercial kits under spin column format and applying a downstream label-free proteomic analysis was published [27]. Therefore, in this context, the aim of our work was to compare five different commercial methods proposed under disposable spin column format, involving reasonable processing time and costs and tested using the same serum pool originated from healthy and diseased individuals. The qualitative and quantitative results obtained for the five conditions were compared together as well as with those obtained with the raw serum analysis. The results are discussed in the context of the application of the prefractionation methods on serum before a clinical biomarker discovery study by label-free proteomics. It can be viewed as a tool to select the most convenient method to apply on clinical sera before proteomic analysis.

2. Materials and methods

2.1. Composition of the sample serum pool

All the serum samples of patients were collected with signed informed consent and with the ethic approval of our university hospital. Three healthy subjects, five patients with colorectal

cancer, five Crohn’s disease and three ulcerative colitis patient samples were selected to prepare a pool of 5 mL. The serum pool aliquots of each patient were stored for several years at -80°C (maximum 5 years) and thawed on ice prior to mixing. This raw serum pool (RS* pool) was tested for total protein quantitation using the RC DC kit (BioRad, Inc., Hercules, CA, USA). Aliquots of suitable volumes were prepared for each protein depletion kit applications and stored at -80°C until further use.

2.2. Prefractionation of the RS* pool

The methods of depletion tested were: ProteoPrep® 20 Plasma kit (Sigma St. Louis, USA), under the spin column format, the ProteoMiner™ (BioRad, Inc., Hercules, CA, USA) designed for 200 μL of sample volume, IgG & Albumin Spin Trap column (GE Healthcare, USA) and WGA Glycoprotein isolation kit (Thermo-Fisher Scientific Inc., USA) with or without prior IgG & Albumin Depletion Spin Trap (GE Healthcare). Fig. 1 summarizes the workflow with the main sample processing steps applied to the RS* pool before performing the proteomic analysis.

2.2.1. ProteoMiner™ kit (BioRad Inc., Hercules, CA, USA)

The small capacity kit (designed for 200 μL of sample) was used for the three technical replicates which were run in parallel. In brief, a 700 μL volume aliquot of the RS* pool was thawed on ice. After centrifugation, $3 \times 200 \mu\text{L}$ of RS* pool were treated in parallel, as recommended by the manufacturer, and resulted in three successive elution steps per 200 μL of RS* treated. All three eluted fractions were pooled together per replicate and stored at -80°C .

2.2.2. ProteoPrep® 20 Plasma kit (Sigma, St. Louis, USA)

This kit was applied as recommended by the manufacturer. We performed three replicates of process consecutively on the same column (and on the same day). Briefly, for each process replicate

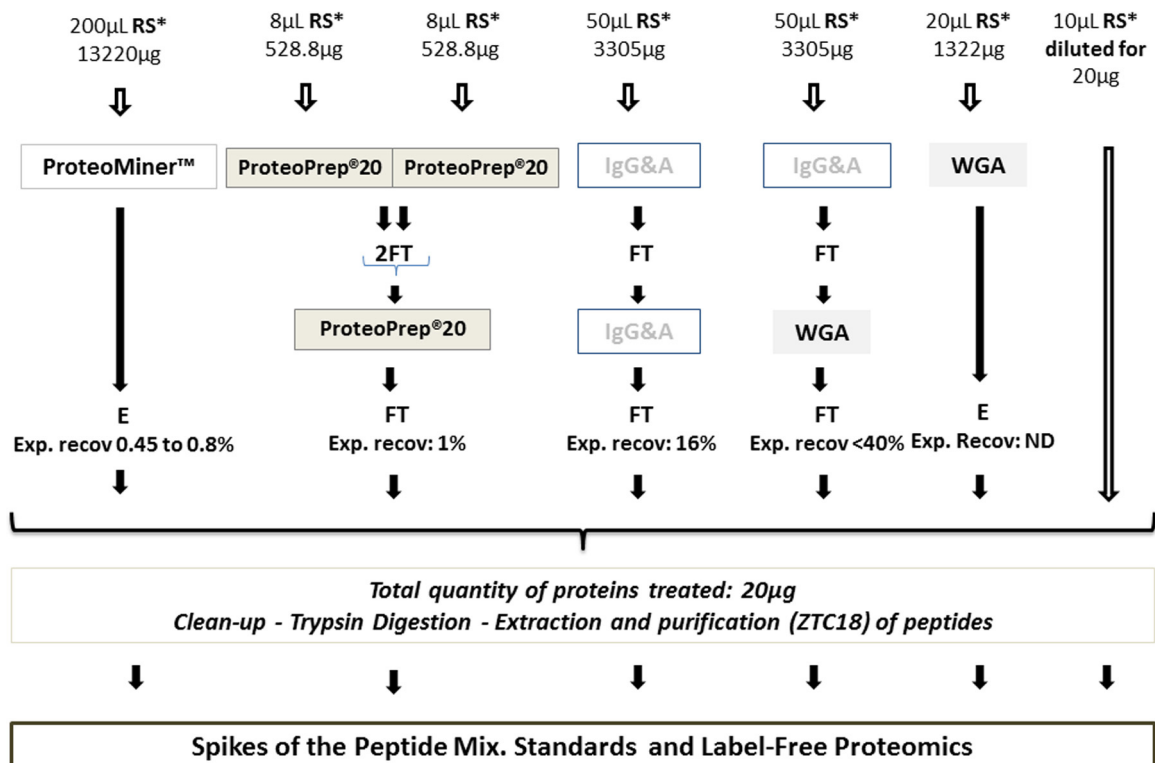


Fig. 1. Experimental workflow detailing the commercial prefractionation and other processing steps applied to the RS* pool.

one aliquot of RS* pool was thawed on ice and 24 μ L were diluted in 300 μ L of binding buffer provided in the kit. The depletion was applied twice on 100 μ L of filtered and diluted spiked serum pool. Afterwards, the two depleted fractions collected were pooled together and concentrated to 100 μ L using the VivaSpin (cut-off 5 kDa) included in the kit. The concentrated pool was applied on column for the second depletion cycle and the final depleted fraction obtained (which was subjected on total to two cycles of affinity depletion) was further stored at -80°C . Reconditioning of the affinity column kit was performed according to the manufacturer recommendation.

2.2.3. Albumin & IgG Spin Trap column (GE Healthcare, UK Ltd., UK)

An aliquot of RS* pool was thawed on ice and 355 μ L of RS* were diluted with the provided kit binding buffer, to allow for six replicates (50 μ L each) applications. These 6 depletions were run in parallel using the protocol recommended by the manufacturer. However, application of the IgG and albumin affinity depletion was applied twice consecutively for every sample replicate processed. The three depleted fractions were stored at -80°C until further use and the three other ones were stored at 4°C and processed using WGA glycoprotein enrichment on the same day.

2.2.4. WGA Glycoprotein isolation kit (Thermo-Fisher Scientific, Inc., USA)

2.2.4.1. On RS* pool. The three technical replicates performed on the spiked serum pool were run in parallel. An RS* pool aliquot was thawed on ice and centrifuged. Only 20 μ L per replicate were used for the kit application as recommended by the manufacturer, using

200 μ g of WGA-lectin beads per replicate. The two steps of elution were performed using 200 μ L elution buffer provided in the kit and pooled together before storage at -80°C .

2.2.4.2. Albumin and IgG depleted RS* pool. The three technical replicates of WGA-lectin glycoprotein enrichment performed after IgG & albumin depletion of the spiked serum pool were run in parallel. The flow-through ($V = 500 \mu\text{L}$) of Albumin & IgG depletion Spin Trap was supplemented with 150 μ L of binding buffer. The next steps of the process were done as recommended by the manufacturer with the two final elutions which were performed using 200 μ L elution buffer provided in the kit. The eluted fractions were all pooled together before storage at -80°C .

2.3. General experiment strategy considerations

In order to ensure standardization of the digestion and peptides conditioning at the end of each preparation process, a necessarily constant quantity of material ($=20 \mu\text{g}$), was treated. (see Fig. 1). All the steps of protein clean-up, reduction-alkylation and the final trypsin digestion, were performed in parallel with a standardized protocol. No other quality control of trypsin digestion was applied and run in parallel of the samples to simplify the QC of the experiment. Indeed we considered that the RS* pool replicates themselves were the QCs of digestion, peptide extraction and MS^E data acquisition and analysis. However, an external QC composed of entire proteins at known relative quantity ratio and known concentrations can be run in parallel of the samples as external digestion QC, but not involving the aspect of matrix of high complexity as serum.

Table 1
Qualitative and quantitative results summary.

	Characteristics	ProteomeMiner™	ProteoPrep® 20	IgG & albumin	IgG & albumin + WGA	WGA	RS*
Sample data before MS analysis	Total starting protein quantity (RCDC on serum pool) (μg)	13220	1057	3305	3305	1322	20
	Volume of RS* pool treated (μL)	200	16	50	50	20	0.3031 ^a
	Recovery of total proteins after prefractionation (mean of 3 replicates \pm SD, expressed in percent of total starting quantity treated) (%)	2.3 ± 0.34	1.91 ± 0.17	24.47 ± 4.93	7.58 ± 0.52	32.91 ± 3.21	NR
PLGS 2.5 ion count tables derived data	Theoretical expected protein quantity recovery (percentage of starting protein quantity) (%)	$0.4-0.7^b$	1	16	<40	NC	100
	Manipulation time for kit application (h)	4	3	3	5.5	2.5	0.1
	Quantity of proteins experimentally identified and quantified by 2D-nanoUPLC-MS ^E run (mean of 3 process replicates \pm SD)(ng)	805.29 ± 113.73	1450.43 ± 241.89	1652.35 ± 315.88	653.12 ± 24.78	686.7 ± 75.15	721.1 ± 141.1
	Number of proteins identified in the 3 replicates of process (3R3) (number of proteins identified and quantified in 3R3)	130 (124)	191 (178)	134 (122)	127 (117)	130 (128)	128 (126)
	Number of proteins identified and quantified in 3R3 with individual Qty SD $\leq 50\%$; Qty SD $\leq 30\%$	97; 70	127; 83	100; 88	98; 89	31; 21	93; 74
	Percentage increase obtained compared to RS* pool for all proteins; for proteins with Qty SD $\leq 50\%$; with Qty SD $\leq 30\%$ (number of proteins gained compared to RS* pool) (%)	$-1.5; 4.3; -5.4$	$49.2; 36.6; 12.2$	$4.7; 7.5; 18.9$	$-1; 5.4; 20.3$	$1.5; -66.6; -71.6$	NR
	Number of proteins identified and quantified in 3R3 whatever Qty SD, reported to the mean protein quantity detected (number of prot/ng)	0.154	0.123	0.074	0.179	0.186	0.175
	Number of unique proteins quantified in all 3 replicates of process (with filter on Qty SD $\leq 50\%$; $\leq 30\%$)	6 (8; 7)	38 (31; 16)	1 (2; 1)	1 (7; 4)	3 (0; 0)	7 (4; 4)
	The concentration dynamic range of proteins identified and quantified in 3R3 ^c (with logarithmic transformation)	3.10	3.86	3.52	3.60	3.30	3.78

NC, not communicated by manufacturer, NR, not relevant.

^a The exact volum treated was 10 μ L of diluted RS* pool in PBS such as 20 μ g of protein were treated and correspond to only 0.303 μ L of undiluted RS* pool.

^b Data deduced from previous tests performed in the laboratory (data not shown).

^c Concentration dynamic range is the ratio between the highest and the lowest quantities of the proteins identified and quantified in all three replicates (a mean is calculated using the three replicates of process performed for both the highest and the lowest abundant proteins obtained); this ratio value was subjected to base 10 logarithmic transformation.

We performed some of the depletions twice consecutively on the same starting material (for IgG & albumin depletion and ProteoPrep[®] 20 kit) to get optimal results. Despite the application of some depletion cycles twice, the time necessary for all steps completion stayed inferior to 6 h per sample replicate (see Table 1). Moreover, these methods at the exception of the ProteoPrep[®] 20 could be run in parallel reducing again the required sample preprocessing time and in theory their respective repeatability variation.

2.4. Protein digestion and preparation for label-free proteomics

After quantitation using the RC DC kit (BioRad, Inc., Hercules, CA, USA), an aliquot of each fractionated sample replicate and the diluted RS* pool (20 µg of total proteins) were precipitated in parallel using 2D-Clean up kit (GE Healthcare, UK). The samples were all reduced and alkylated as previously described and digested using porcine recombinant trypsin (Roche, Ltd., Switzerland) [28].

Each protein digest was placed in aqueous 0.1% trifluoroacetic acid final solution and further purified using C18 ZipTip high capacity (Merck Millipore, Billerica, MA, USA). Protein digests were speed-vacuum dried and reconstituted using 100 mM ammonium formate solution adjusted to pH 10 at a concentration of 278 ng/µL. Each injected sample was spiked with a commercial mixture of protein digest standards originated from non-human biological material: the MassPREP[™] digestion standards (Waters, Corp., Milford, USA) at 150 fmol of ADH per sample process replicate. This commercial standard consists of two standard mixtures (MPDS Mix 1 and MPDS Mix 2) containing protein digests of yeast alcohol dehydrogenase (ADH), rabbit glycogen phosphorylase b, bovine serum albumin, and yeast enolase present at known protein quantities; allowing therefore to check for relative quantitation of the samples spiked. Sample injection replicates order on 2D-nanoUPLC-ESI-MS^E (Waters, Corp., Milford, USA) was randomized and 2.5 µg of total protein digest (in 9 µL) were injected per sample replicate.

2.5. 2D-nanoUPLC-ESI-MS^E system configuration and settings

All eighteen samples (three × five pre-fractionated digested samples and three × RS* pool digested), were injected onto the 2D-nanoAquity UPLC (Waters, Corp., Milford, USA) coupled online with the Q-TOF Synapt HDMS[™] G2 system (Waters, Corp., Milford, USA) using ion mobility as supplementary separation. The configuration of the 2D-nanoUPLC system was a reversed phase pH 10–reversed phase pH 3 based two dimension separation. The first dimension separation was made on an X-Bridge BEH C18 5 µm column (300 µm × 50 mm). The trap column Symmetry C18 5 µm (180 µm × 20 mm) and analytical column BEH C18 1.7 µm (75 µm × 250 mm) (Waters, Corp., Milford, USA) were used after an online dilution to lower pH values (minimum been pH 3). The samples were loaded at 2 µL/min (20 mM ammonium formate solution adjusted to pH 10) on the first column and subsequently eluted in five steps (10, 14, 16, 20 and 65% acetonitrile). Each eluted fraction was desalted on the trap column after a ten times online dilution to pH 3 and subsequently separated on the analytical column; flow rate 250 nL/min, solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile), linear gradient 0 min, 97% A; 90 min, 65% A. (The total run time of analysis was 5 × 120 = 600 min).

The mass spectrometer was operated in positive ion mode and the data acquisition was performed in the 50–1500 *m/z* range, with a scan time of 0.6 s and with collision energy voltages set in independent alternative scanning (MS^E) mode. The IMS cell pressure was set at 2.5 mbar, the variable IMS wave velocity

ranged from 850 m/s to 1200 m/s and the wave height was 40 V. A lock mass correction was done using [Glu1]-fibrinopeptide B ([M + 2H]²⁺: 785.84206 *m/z*) (Waters Corp., Milford, USA).

2.6. Processing of raw data using ProteinLynx GlobalServer (PLGS 2.5, Waters Corp., Milford, USA) and analysis of ion account tables

Raw data processing (deconvolution and deisotoping), protein identification and relative quantification were all performed using ProteinLynx Global SERVER (PLGS) v2.5 (Waters Corp., Milford, USA). The processing parameters were set as follows: the MS TOF resolution and the chromatographic peak width were both set to automatic, the low-/elevated-energy detection threshold was set to 150/15 counts, respectively, the identification intensity threshold was 500 counts and the lock mass window at 785.84206 *m/z* was set to 0.30 Da.

For the identification of the proteins, the UniProt human database (UniProt release 2011_12-December 14th, 2011) implemented with the sequences of the non-human protein standards spiked as proteins digests before injection in raw serum (accession #: P00489, P00924, P02769, P00330), was used. The searched parameters used were as follows: carbamidomethylation (C) and oxidation (M) of peptides as fixed and variable modifications respectively, two possible trypsin missed cleavages were allowed, minimum fragment ion matches per peptide of three and minimum fragment ion matches per protein of seven, minimum two peptides matches per protein and finally a protein false discovery rate (FDR) for protein was set at maximum 4% (which is equivalent to a 1% FDR at the peptide level).

Protein and peptides ion account tables were generated during protein identification. The protein quantitation was then based on calibration using the response factor of the ADH spiked. The normalization on ADH was also performed using the PLGS vs 2.5 Expression module to control technical variability. The technical variability was evaluated using comparison between expected and obtained protein ratios for the different sample standard Mix MassPREP[™] digestion standards Mix 1 or 2 (Waters Corp., Milford, USA). These were calculated from the expression analysis tables generated using P00330-ADH normalization to verify that the process was technically sound and that the variability of this experiment satisfied the one of this label free technology (maximum 30%) [29]. Ion account tables were used to extract data concerning the proteins identified and quantified in each replicate done for each samples. The method of quantitation used for ion account table generation by PLGS vs 2.5, is the “Top 3” [30]. To enable comparison of different pre-fractionation methods applied to the spiked raw serum, we expressed the quantity of each protein as the percentage of total protein quantity obtained per replicate run. We calculated each mean protein quantity value based on the three replicate values for the total protein quantity and also for each individual protein quantity. We also calculated the protein concentrations equivalent to the initial quantity (and volume) of RS* pool treated. This one is referred as the “calculated raw serum equivalent concentration of protein” and is expressed in µg/mL. It was obtained by dividing by 100 each protein relative quantity (obtained using the 3 replicates of process and expressed in percents), and by multiplying the quotient obtained by a factor. This factor was defined as the ratio between the quantity of total protein recovered after the pre-fractionation and the quantity analyzed on column. For example, for ProteoPrep[®] 20 sample: 20.1887 µg/2.5 µg = 8.0755. This last product was further multiplied by the mean total protein quantity (obtained with the 3 process replicates and expressed in µg) and divided by the number of mL of raw serum treated initially (for ProteoPrep[®] 20: 0.016 mL). The ratio obtained was therefore different for each kit. For example, for the ProteoPrep[®] 20, the ratio is 8.075480 as

20.1887 μg was recovered and that only 2.5 μg was loaded on column. For RS* pool, this ratio is 8 (20 μg /2.5 μg). These data were used to compare the 5 prefractionation methods together and with the RS* pool results.

The RSD was obtained for the total protein quantity recovery calculated using each process replicates (=relative Qty SD/mean relative Qty \times 100, expressed in %).

All the results data were treated using the Excel tabler.

The *Expression Analysis* module was used to perform differential analysis using the three technical replicates done per depletion/enrichment kit, performing exclusively pairwise comparisons. Expression analysis was used to verify the spike protein digest of the standard Mix MassPREP™ digestion standards Mix 1 and 2 ratio. Normalization was performed on the ADH (P00330). This analysis was done only considering the proteins identified and quantified in three out of three replicates of process (filter application: 3R3).

3. Results and discussion

3.1. Total protein quantity obtained for the prefractionated samples

The measured protein concentration of the RS* pool was $66.106 \pm 1.890 \mu\text{g}/\mu\text{L}$. The total protein quantities and the corresponding sample volumes treated with each method, the expected theoretical data and the empirical results obtained are summarized in Table 1. The repeatability of each kit applications was evaluated using the total protein quantity measured just prior to sample clean-up and it provided a SD < 5%.

3.2. Proteomic results analysis

3.2.1. Data technical quality controls

The MPDS mix protein ratios were calculated using the quantity of ADH (P00330) spiked amount, while performing expression

analysis to compare process replicates of each method of serum preprocessing. All the Mix MassPREP™ digestion standards proteins showed process replicates SD equal or lower to the expected 30% [31].

3.2.2. Qualitative and quantitative proteomic results

Table 1 summarizes the main results obtained for each method of prefractionation and the RS* pool.

3.2.2.1. Total protein recovery. The total protein recovery obtained after prefractionation and the quantity for each protein identified and quantified are detailed in Table 1. The quantity of each protein is the mean of the three process replicates \pm standard deviation (SD), obtained for each prefractionation method.

3.2.2.2. Individual protein variability. The variability in protein quantity identified and quantified per process replicates is expressed in ng (see Table 1). The relative standard variations (RSD) of each sample were lower than 20%. The RS* pool showed in this experiment the highest RSD at 19.57% counting only for clean-up, digestion and peptide resolubilisation/extraction yields and the applied downstream nanoUPLC-MS^E step repeatability. The smallest RSD was obtained with the WGA lectine glycoprotein enrichment applied after IgG & albumin depletion (RSD = 3.79%). The RSD of ProteoMiner™ and ProteoPrep® 20 were 14.12% and 16.65%, respectively. The related product “prot20 column” (Sigma) prefractionation on an off-line FPLC-system combined with label-free LC-MS/MS analysis provided inter-runs CV assays of 30.9% [12]. Moreover, the distribution of each protein RSD obtained for the RS* pool was similar to the one of other pretreated samples (data not shown). Hakimi et al. using plasma sample and spin column format of ProteoMiner™ and ProteoPrep® 20, reported a better reproducibility of process using ProteoMiner™ than using the ProteoPrep® 20 depletion strategy [27]. On the serum pool

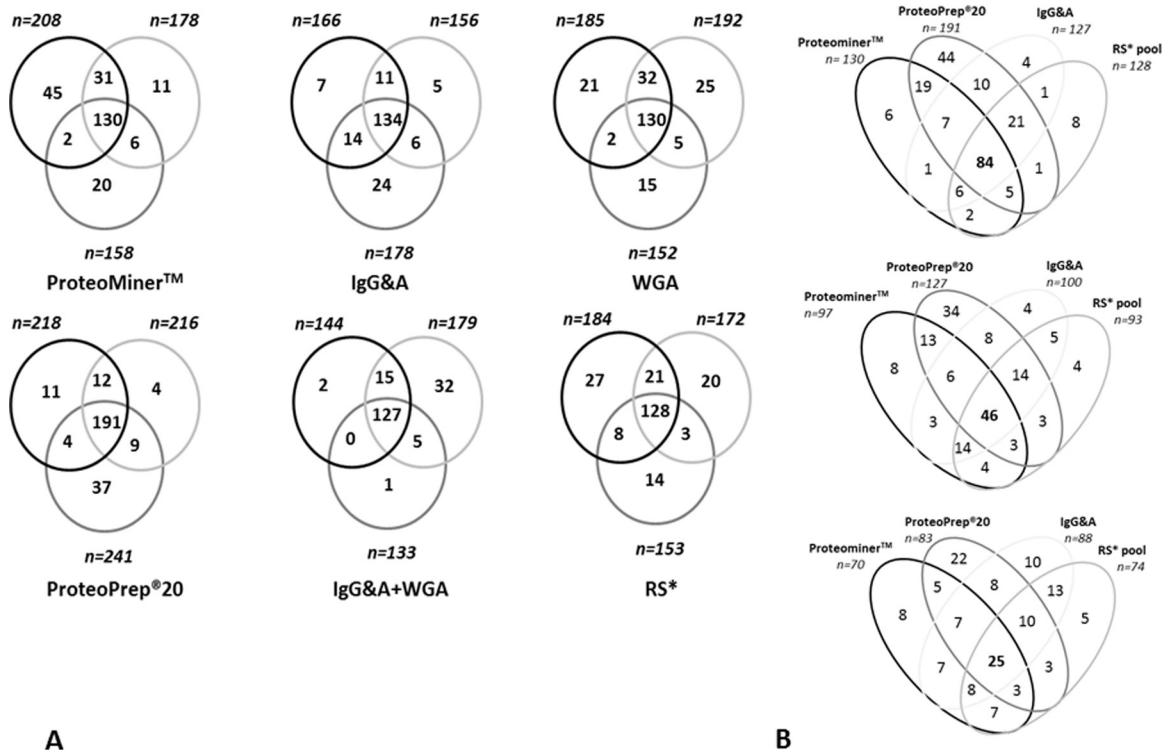


Fig. 2. (A) Venn diagrams with the proteins identified per process replicate of each prefractionation method and the RS* pool, all analyzed by label-free MS^E proteomics. (B) Comparison illustrated by Venn diagrams of the proteins identified in 3R3 for the ProteoPrep® 20, for the ProteoMiner™, IgG & albumin depleted samples and the RS* pool. From top to bottom the Venn analysis results are shown applying, respectively, no, 50% and 30% threshold quantity SD filter on proteins included.

tested in our work, ProteoMiner™ appeared slightly better than ProteoPrep® 20.

3.2.2.3. Number of different proteins identified and quantified

3.2.2.3.1. Data for each method and the RS* pool. Table 1 details the number of proteins identified in three replicates out of three (3R3). The range and the percentage of conserved proteins after the 3R3 filter application shows the qualitative repeatability of each method (=prefractionation process and nanoUPLC–MS^E). This is illustrated Fig. 2A with Venn diagrams built with the number of proteins identified in each replicate of process (spiked proteins excluded). The numbers of proteins quantified and showing a quantity SD higher than 30% (technical variation of the instrumental system) and lower than 50% (arbitrary cut-off value that might be suitable for a discovery phase) are annotated in the Table 1. As a discovery study should include some biological replicates to be consistent and as the variability associated to biological replicates is certainly higher than the technical one, we included an analysis of the data after applying this 50% cut-off for quantity SD (%). This indicated the number of proteins that might be of true value for consistent results analysis after sample preprocessing. The 30% threshold filter corresponding to the technical variability of the instrumental system used in this experiment could be considered as too stringent.

3.2.2.3.2. Comparison of the results obtained after the prefractionations to the RS* Pool. Fig. 2 B shows Venn diagrams comparing three selected methods together and with the RS* pool. The comparison of the six protein lists identified in 3R3, without any quantity SD filter, can be downloaded as Supplementary data Table 1. In the analysis including all the proteins whatever quantity SD obtained, only 71 proteins were found common to the 6 conditions compared. The maximum number of “unique proteins”, identified solely in one prefractionation method, with 3R3 filter application and with a quantity SD lower to the two thresholds are also reported in Table 1. The numbers of “unique proteins” indicate the advantage of one method over the others and mostly over no prefractionation (RS* pool sample). ProteoPrep® 20 provided the highest number of “unique proteins” whatever

quantity SD filter is applied. In addition, ProteoPrep® 20 process enabled a 49% increase in number of proteins identified compared to the RS* pool (in 3R3; no Qty SD filter). But when analyzing the number of proteins with a suitable quantity SD, the gain of 36.6 % obtained for 50% Qty SD dropped to 12.2%, when using the technical 30% threshold. At this level of stringency, the Albumin & IgG Depletion Spin Trap kit or the combination of IgG & albumin depletion prior to WGA enrichment appeared to be slightly superior with a gain of 18.9% and 20.3%, respectively over the RS* pool. The apparent increases observed for some of the prefractionation kits used must be nuanced depending on the stringency of analysis. Altogether, ProteoPrep® 20 enrichment provided the highest advantage, regarding the quantity/volume ratio of raw biological material needed for the kit application and the number of proteins significantly identified and quantified. All the other prefractionation methods tested regarding these criteria did not provided better results than the RS* pool.

3.2.3. The most abundant proteins and their depletion efficiencies

3.2.3.1. The 20 most abundant protein families analysis. Supplementary data Table 2 details all the proteins and protein families targeted by the ProteoPrep® 20 plasma kit (according to information provided by the manufacturer data sheet). The data provided to facilitate comparison of kit efficiencies are the relative quantities (% of the total quantity of protein detected and quantified for each condition). When summing the relative protein quantity percentages of the IgG family members identified (Prot.entry and Prot.Acc # with ** in Supplementary data Table 2), the ProteoPrep® 20 method showed the lowest total relative quantity, which indicates their more efficient depletion with this kit. The calculated protein concentrations, equivalent to raw serum ($\mu\text{g}/\text{mL}$) and the corresponding SD are also communicated for each sample. But, not all the most abundant proteins pointed showed an individual quantity SD < 30%.

3.2.3.2. The efficiency of albumin (HSA) depletion. The method that provided the best results concerning HSA depletion was the combination of IgG & albumin depletion with downstream WGA

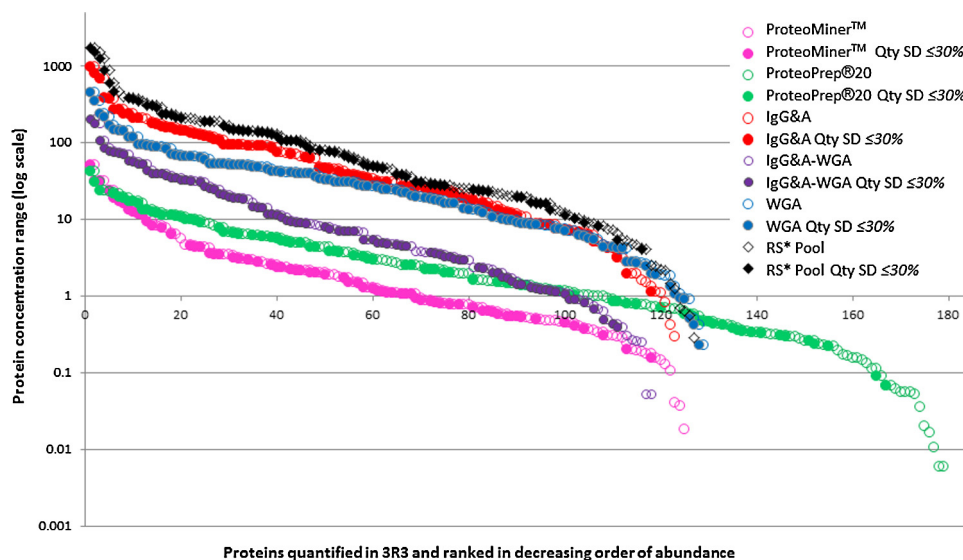


Fig. 3. The six ranges of calculated raw serum equivalent protein concentration obtained for the proteins identified and quantified in 3R3 ($\mu\text{g}/\text{mL}$) for the methods of prefractionation and for the RS* pool sample.

Protein with quantity SD \leq 30% (filled label: ●); Quantity SD > 30% (empty label: ○)

The lowest abundant proteins reliably quantified in each sample (Qty SD \leq 30%) are K1C25 ($0.018 \pm 0.031 \mu\text{g}/\text{mL}$) for ProteoMiner™; ACTH ($5.84 \cdot 10^{-3} \pm 10.13 \cdot 10^{-3} \mu\text{g}/\text{mL}$) for ProteoPrep® 20; K1C28 ($0.297 \pm 0.514 \mu\text{g}/\text{mL}$) for IgG & albumin; K1C28 ($0.052 \pm 0.089 \mu\text{g}/\text{mL}$) for IgG & albumin + WGA; KRT15 ($0.227 \pm 0.394 \mu\text{g}/\text{mL}$) for WGA; HBD ($0.288 \pm 0.500 \mu\text{g}/\text{mL}$) for RS* pool.

enrichment which showed the lowest relative quantity percentage. The second best was WGA enrichment and the worst was ProteoMiner™ with results closer to these of RS* pool. Knowing that HSA is present in serum at approximately 75% of total protein quantity [32], the low percentage in RS* pool (9.15% of HSA quantified) illustrated the compression of the protein concentration dynamic range obtained. This was however expected using this proteomic technology and these settings, potentially involving also suppression effects well known with electrospray ionization (ESI) and the fact that the quantitation of analytes with this technology is concentration dependent [33]. However, one must keep in mind that very highly concentrated proteins (as HSA) and therefore the most abundant ones, could be saturating and not be correctly quantified. Importantly, HSA showed, only for RS*pool, ProteoMiner™ and WGA enrichment, quantity SD lower to 30%.

3.2.4. The protein concentration dynamic ranges and the abilities to measure low-abundance proteins in serum

3.2.4.1. Protein concentration dynamic range. Table 1 provides the dynamic ranges of protein concentrations observed (with base 10 logarithmic transformation) for the six conditions analyzed. Fig. 3 shows for every method tested, the distribution of the proteins identified and quantified and ranked in decreasing order of calculated equivalent to raw serum concentrations (using a base 10 logarithmic scale). The largest distribution in term of protein concentrations was obtained using the ProteoPrep® 20 and second best was the RS* pool itself. The smallest range observed was obtained after equalization with ProteoMiner™ which might be obvious considering the theoretical principle of this kit. This is in opposition with what was reported previously using similar workflows: DIA and label-free strategy [27]. In this work, ProteoMiner™ application on plasma enabled detection of proteins within a 5 orders logarithmic range of protein concentrations. This demonstrates again the shrinking effect of this strategy when applied on serum matrix, as reported previously [21,34]. But also highlight the difference of results obtained when using plasma or serum. Fig. 3 shows that the maximum gain in protein identifications number (49%) obtained with ProteoPrep® 20, were proteins lying within medium values of the protein relative quantification range. The ProteoMiner™ curve obtained was not as flat as could be expected after equalization. But the results obtained for the 6 conditions (when considering only individual proteins with quantity SD ≤ 30%, (illustrated using the filed label on the graph), allowed the same conclusions, but with a dynamic range 0.5–1 order shorter. Supplementary data Table 3 illustrates also the deformation in the proteins concentration dynamic range observed for the six conditions compared. Some proteins were selected at different representative relative quantities and distributed within the range observed in the RS* pool. Their corresponding calculated equivalent to raw serum protein concentrations are provided, as well as known concentrations measured in serum [35], plasma concentration observed [1,36,37] or estimated [38]. The ranking of abundances were affected when using prefractionation and some proteins showed quantity below the lower limit of quantitation (LLOQ) of the method. The deformation observed might be driven by compression effects and also by some truncations in the concentration range due to the depletion of some specific proteins, also visible in Fig. 3. For example the vitamin D binding protein (VTDP- P02774), not related to cell structure but to the active analyte of vitamin D transport, was not expected to be within the low concentration range in serum. Its concentration ranges from 193 to 4350 mg/L in serum of healthy individuals (all phenotypes considered) [35]. But, some values obtained for some

of the kits tested (see Supplementary data Table 3) are very different and cannot all be reliably considered (because of Qty SD > 30 %). For example in ProteoMiner™, it is detected at only 0.6% of the value found for the RS* pool sample. But binding of proteins with ProteoMiner™ has been shown to be driven by non specific hydrophobicity rather than by the hexapeptide affinity. In this mechanism the hydrophobicity of each protein present in the complex mixture drives their own respective binding and depletion [4]. However, abundant hydrophilic proteins as IgG or apolipoprotein A1 (ApoA1) were merely depleted efficiently with this method, which was also observed in this serum data set (see Supplementary data Table 2). Importantly, some proteins were quantified for some samples only with a quantity SD higher to 30 or 50% and might or should therefore not be considered as reliably quantified.

3.2.4.2. The ability to reach low-abundance proteins. To evaluate if the prefractionations tested enabled the analysis of proteins present at low concentrations in serum (the deep proteome), one should compare the proteins quantified as lowest ones in each sample to their level in serum determined by other techniques. But most of these serum protein levels are not available in literature, despite that valid reference measurements or estimations might be available for plasma [36]. Therefore, no reference could be used to compare the gain in low-abundance serum proteins reached by these kits. However, we analyzed for each process replicate, the lowest quantity detected. All the kits provided similar results with no significant difference (Mann–Withney test, P value > 0.05) to these obtained with the RS* pool sample (lowest single replicate value = 0.010 ng, with a mean value over the triplicates of 0.022 ± 0.027 ng). This was due to the main limitations of the experimental design, where a constant quantity of total protein digest was loaded on column, in addition to the lower limit of quantification (LLOQ) of the technology itself. However, when considering the results expressed in terms of calculated raw serum equivalent protein concentrations, and with quantity SD < 30% filter, the lowest values were found for Dermicin (P81605). Its mean concentration is $0.067 \mu\text{g/mL} \pm 21\%$ in the ProteoPrep® 20 sample. The ratio between the lowest protein quantities obtained in the RS* pool and the ProteoPrep® 20 is 49 (61 when considering the proteins with quantity SD ≤ 30%). The three methods enabling to measure proteins present at lower equivalent raw serum concentrations compared to RS* pool were IgG & albumin depletion coupled with WGA enrichment, ProteoMiner™ and ProteoPrep® 20, showing itself the lowest value. Hence, despite that the comparison of our results with literature serum references was not always possible, we could conclude that in our experiment and data set, ProteoPrep® 20 enabled the lowest calculated equivalent raw serum concentrations of protein, with 49–61 times lower values than these obtained with the RS* pool.

3.2.5. Comparison of the results of ProteoPrep® 20 treated sample to reference concentrations of proteins

The list of proteins identified and quantified in the ProteoPrep® 20 sample is available as Supplementary data Table 4. This table provides the quantity of each protein and SD (both expressed in percentage of total protein quantity detected) and each protein Quantity SD (in %). The calculated ProteoPrep® 20 depleted serum concentration ($\mu\text{g/mL}$) of proteins equivalent to the initial raw serum and also, some known reference concentrations of these proteins taken from the “high confidence plasma proteome reference set” are reported [38]. These concentration references in plasma or serum are provided for ranking comparison of the proteins found in a non depleted serum or plasma and identified after ProteoPrep® 20 serum depletion. Regarding the proteins

listed (191 protein identified in 3R3 and 178 quantified) only 38 proteins (in bold) were identified only after using this depletion method (=“unique proteins”). Compared to the 1929 non canonical sequences reported by Farrah et al., 45 proteins were identified only in our analysis. These are not all related to the complement activation and the coagulation process as would be expected in serum compared to plasma. These proteins might also be related to the disease state of the patient sera included in our RS* pool.

3.3. General conclusions

The five methods tested together with label-free proteomic analysis on a 2D-nanoUPLC-MS^E Synapt G2 system showed acceptable repeatabilities, similar to the one obtained for the RS* pool sample analysis. All five methods showed equivalent performances regarding total protein quantity recovery. Surprisingly, the RS* pool already provided very interesting results with this system settings, which illustrates its power in terms of peptides and ions separations and proteins identifications, as well as relative quantification. But, some of the prefractionations tested did not add much information over the raw serum analysis. This is worth knowing before spending time, energy and resources on precious clinical samples prefractionation. In summary, in our hands, the ProteoPrep[®] 20 which is the only fully re-usable kit present the following advantages. First, it requires the lowest volume or quantity of precious clinical sera. Second it enables the maximum number of protein identifications increase (49%) compared to the RS* pool with the highest number of unique proteins. However other depletion protocols might even be more valuable if very stringent Qty SD threshold have to be applied. These involves IgG & albumin depletion strategy. Nevertheless, the third main advantage of the ProteoPrep[®] 20 relies on the possibility to reach up to 61 times lower protein concentration values compared to non depleted serum which are low abundance protein not always detectable for other kits.

But this depletion kit is not perfect as it is expensive and time consuming (two successive cycle of depletions were necessary to reach the purity obtained in this work). This antibody affinity capture spin column is a reusable format and despite the warranty that one hundred applications might be done without losing depletion efficiency, bias in sample processing for a clinical sample set due to the order of samples prefractionation could still occur. Hence, the order of samples should always be randomized, shuffling diseases and controls (and likely QC) to minimize the potential effect as such a bias risk. Moreover, potential cross-contaminations remain possible, even if careful reconditioning of the spin column is performed between the depletion cycles.

Altogether, albeit not perfect, ProteoPrep[®] 20 appeared in this experimental design and data set to be a suitable compromise for application as off-line serum samples prefractionation strategy. The results of a differential label-free proteomic analysis for clinical biomarkers discovery phase would profit from this type of prefractionation as long as results are analyzed knowing the technical repeatability of the global strategy.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We thank the GIGA Proteomic Facility from ULg, involved in shotgun proteomic analyses and Mrs. Trzpiot for technical assistance.

This research was supported by the “Centre Anti-Cancereux”, CHU de Liège and the FEDER action: “Fédération Wallonie Bruxelles”.

This work was also made possible thanks to the support of MSD. NS was a FNRS logistic collaborator.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.euprot.2015.07.009>.

References

- [1] M. Polanski, N.L. Anderson, A list of candidate cancer biomarkers for targeted proteomics, *Biomark. Insights* 1 (2007) 1–48.
- [2] M. Pernemalm, R. Lewensohn, J. Lehtio, Affinity prefractionation for MS-based plasma proteomics, *Proteomics* 9 (2009) 1420–1427.
- [3] N.L. Anderson, N.G. Anderson, The human plasma proteome: history, character, and diagnostic prospects, *Mol. Cell. Proteomics* 1 (2002) 845–867.
- [4] E.M. Keidel, D. Ribitsch, F. Lottspeich, Equalizer technology—equal rights for disparate beads, *Proteomics* 10 (2010) 2089–2098.
- [5] B. Domon, R. Aebersold, Options and considerations when selecting a quantitative proteomics strategy, *Nat. Biotechnol.* 28 (2010) 710–721.
- [6] J.R. Whiteaker, H. Zhang, J.K. Eng, R. Fang, B.D. Piening, L.C. Feng, et al., Head-to-head comparison of serum fractionation techniques, *J. Proteome Res.* 6 (2007) 828–836.
- [7] S.J. Geromanos, J.P. Vissers, J.C. Silva, C.A. Dorschel, G.Z. Li, M.V. Gorenstein, et al., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS, *Proteomics* 9 (2009) 1683–1695.
- [8] G.Z. Li, J.P. Vissers, J.C. Silva, D. Golick, M.V. Gorenstein, S.J. Geromanos, Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures, *Proteomics* 9 (2009) 1696–1719.
- [9] Y.S. Choi, Reaching for the deep proteome: recent nano liquid chromatography coupled with tandem mass spectrometry-based studies on the deep proteome, *Arch. Pharm. Res.* 35 (2012) 1861–1870.
- [10] Q. Zhang, V. Faca, S. Hanash, Mining the plasma proteome for disease applications across seven logs of protein abundance, *J. Proteome Res.* 10 (2011) 46–50.
- [11] L. Huang, X. Fang, Immunoaffinity fractionation of plasma proteins by chicken IgY antibodies, *Methods Mol. Biol.* 425 (2008) 41–51.
- [12] M.P. Smith, S.L. Wood, A. Zougman, J.T. Ho, J. Peng, D. Jackson, et al., A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses, *Proteomics* 11 (2011) 2222–2235.
- [13] C. Tu, P.A. Rudnick, M.Y. Martinez, K.L. Cheek, S.E. Stein, R.J. Slebos, et al., Depletion of abundant plasma proteins and limitations of plasma proteomics, *J. Proteome Res.* 9 (2010) 4982–4991.
- [14] W.J. Qian, D.T. Kaleta, B.O. Petritis, H. Jiang, T. Liu, X. Zhang, et al., Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy, *Mol. Cell. Proteomics* 7 (2008) 1963–1973.
- [15] T. Shi, J.Y. Zhou, M.A. Gritsenko, M. Hossain, D.G. Camp 2nd, R.D. Smith, et al., IgY14 and SuperMix immunoaffinity separations coupled with liquid chromatography-mass spectrometry for human plasma proteomics biomarker discovery, *Methods* 56 (2012) 246–253.
- [16] P.G. Righetti, E. Boschetti, L. Lomas, A. Citterio, Protein equalizer technology: the quest for a democratic proteome, *Proteomics* 6 (2006) 3980–3992.
- [17] P.G. Righetti, E. Boschetti, The ProteoMiner and the FortyNiners: searching for gold nuggets in the proteomic arena, *Mass Spectrom. Rev.* 27 (2008) 596–608.
- [18] C. Sihlbom, I. Kanmert, H. Bahr, P. Davidsson, Evaluation of the combination of bead technology with SELDI-TOF-MS and 2-D DIGE for detection of plasma proteins, *J. Proteome Res.* 7 (2008) 4191–4198.
- [19] L. Guerrier, S. Claverol, F. Fortis, S. Rinalducci, A.M. Timperio, P. Antonioli, et al., Exploring the platelet proteome via combinatorial, hexapeptide ligand libraries, *J. Proteome Res.* 6 (2007) 4290–4303.
- [20] R.C. Dwivedi, O.V. Krokhn, J.P. Cortens, J.A. Wilkins, Assessment of the reproducibility of random hexapeptide peptide library-based protein normalization, *J. Proteome Res.* (2010) 1144–1149.
- [21] J. Rivers, C. Hughes, T. McKenna, Y. Woolerton, J.P. Vissers, J.I. Langridge, et al., Asymmetric proteome equalization of the skeletal muscle proteome using a combinatorial hexapeptide library, *PLoS One* 6 (2011) e28902.
- [22] J.E. Bandow, Comparison of protein enrichment strategies for proteome analysis of plasma, *Proteomics* 10 (2010) 1416–1425.
- [23] M. De Bock, D. de Seny, M.A. Meuwis, A.C. Servais, T.Q. Minh, J. Closset, et al., Comparison of three methods for fractionation and enrichment of low molecular weight proteins for SELDI-TOF-MS differential analysis, *Talanta* 82 (2010) 245–254.

- [24] L.A. Echan, H.Y. Tang, N. Ali-Khan, K. Lee, D.W. Speicher, Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma, *Proteomics* 5 (2005) 3292–3303.
- [25] Z. Cao, H.Y. Tang, H. Wang, Q. Liu, D.W. Speicher, Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes, *J. Proteome Res.* (2012) .
- [26] D.C. Liebler, A.J. Ham, Spin filter-based sample preparation for shotgun proteomics, *Nat. Methods* 6 (2009) 785.
- [27] A. Hakimi, J. Auluck, G.D. Jones, L.L. Ng, D.J. Jones, Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS, *Proteomics* (2013) .
- [28] M. Fleron, Y. Greffe, D. Musmeci, A.C. Massart, V. Hennequiere, G. Mazzuchelli, et al., Novel post-digest isotope coded protein labeling method for phospho- and glycoproteome analysis, *J. Proteomics* 73 (2010) 1986–2005.
- [29] J.C. Silva, R. Denny, C. Dorschel, M.V. Gorenstein, G.Z. Li, K. Richardson, et al., Simultaneous qualitative and quantitative analysis of the *E. coli* proteome: a sweet tale, *Mol. Cell. Proteomics* 5 (2006) 589–607.
- [30] J.C. Silva, M.V. Gorenstein, G.Z. Li, J.P. Vissers, S.J. Geromanos, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition, *Mol. Cell. Proteomics* 5 (2006) 144–156.
- [31] J.P. Vissers, J.I. Langridge, J.M. Aerts, Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease, *Mol. Cell. Proteomics* 6 (2007) 755–766.
- [32] L. Urbas, P. Brne, B. Gabor, M. Barut, M. Strlic, T.C. Petric, et al., Depletion of high-abundance proteins from human plasma using a combination of an affinity and pseudo-affinity column, *J. Chromatogr. A* 1216 (2009) 2689–2694.
- [33] K. Tang, et al., Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* 15 (10) (2004) 1416–1423.
- [34] L. Guerrier, P.G. Righetti, E. Boschetti, Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library, *Nat. Protoc.* 3 (2008) 883–890.
- [35] D. Blanton, Z. Han, L. Bierschenk, M.V. Linga-Reddy, H. Wang, M. Clare-Salzler, et al., Reduced serum vitamin D-binding protein levels are associated with type 1 diabetes, *Diabetes* 60 (2011) 2566–2570.
- [36] G.L. Hortin, D. Sviridov, N.L. Anderson, High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance, *Clin. Chem.* 54 (2008) 1608–1616.
- [37] B.B. Haab, B.H. Geierstanger, G. Michailidis, F. Vitzthum, S. Forrester, R. Okon, et al., Immunoassay and antibody microarray analysis of the HUPO plasma proteome project reference specimens: systematic variation between sample types and calibration of mass spectrometry data, *Proteomics* 5 (2005) 3278–3291.
- [38] T. Farrah, E.W. Deutsch, G.S. Omenn, D.S. Campbell, Z. Sun, J.A. Bletz, et al., A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas, *Mol. Cell. Proteomics* 10 (2011) M110.6353.