SciVerse ScienceDirect

journal homepage: http://www.elsevier.com/locate/euprot

Proteomics of human plasma: A critical comparison of analytical workflows in terms of effort, throughput and outcome

Loïc Dayon[a,*], *Martin Kussmann*[a,b,c]

[a] *Molecular Biomarkers Core, Nestlé Institute of Health Sciences, Lausanne, Switzerland*
[b] *Faculty of Life Sciences, Ecole Polytechnique Fédérale, Lausanne, Switzerland*
[c] *Faculty of Science, Aarhus University, Aarhus, Denmark*

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form 5 July 2013
Accepted 8 August 2013

Keywords:
Blood
Biomarker
Diagnosis
Liquid chromatography
Mass spectrometry
Shotgun proteomics

ABSTRACT

In this study, several workflows to analyze human plasma proteins with RP-LC MS/MS are evaluated. The impact of depletion of abundant proteins on the plasma proteome coverage was assessed together with the duration of RP-LC separation. An additional upstream liquid-based fractionation was evaluated. The applicability and feasibility of these technologies in large-scale clinical studies with respect to effort, throughput, and outcome are discussed. Label-free and isobaric tagging-based quantifications are examined in this perspective. We demonstrate that, despite the great improvement of proteomic technologies, significant trade-offs between effort and yield are still challenging the discovery of protein biomarkers in blood plasma.

1. Introduction

Because of its minimally invasive accessibility and its ready availability, blood is the most preferred and used human body fluid to be measured in routine clinical practice. Moreover, blood perfuses all body tissues and contains, for instance, proteins secreted, shed or released from cells and tissues. Its composition is therefore relevant as an indicator of the overall physiology of an individual [1]. The clinical deployment of human plasma protein markers for disease diagnosis, as well as health/disease prognosis and monitoring is in great demand and requires increased translation of candidates. By contrast, the identification of novel biomarkers in human plasma is highly challenging and the rate of transfer of candidates into the clinics has lagged behind expectations [2]. In protein biomarker discovery projects, body fluids and/or tissues are usually first used to generate lists of biomarker candidates, classically by comparison of case and control samples, more recently also in longitudinal studies, in which each "case" subjects is its own control. To refine the list of putative biomarkers, an analytical verification step is then performed using targeted approaches in an increasing number of plasma samples (typically tens to hundreds). Ultimately, clinical validation of these analytically verified candidates in blood is performed in large cohorts to demonstrate the

* *Corresponding author at*: Molecular Biomarkers Core, Nestlé Institute of Health Sciences SA, Campus EPFL, Quartier de l'innovation, bâtiment H, 1015 Lausanne, Switzerland. Tel.: +41 21 632 6114.
E-mail addresses: loic.dayon@rd.nestle.com, loic.dayon@gmail.com (L. Dayon).

2212-9685
http://dx.doi.org/10.1016/j.euprot.2013.08.001

utility, specificity and sensitivity of a biomarker or a panel of biomarkers and to show their clinical relevance as prognostic or diagnostic tools. The verification and validation steps often require the development of assays, such as enzyme-linked immunosorbent assays (ELISA) or selected-reaction monitoring (SRM) mass spectrometry (MS)-based assays that require time, investment, and expertise for execution and, possibly, also upstream development [3]. The opportunity to perform the first-stage screening in a higher number of samples may help increase the statistical significance of the discovery findings and reduce the number of "pilot" assays to be developed, which again requires substantial resources. Especially in view of the latter aspect, analytical throughput of the proteomic discovery workflow is an important aspect to be considered [4].

The proteomic analysis of blood plasma with MS is a challenging task. The high complexity and wide dynamic range of proteins as well as the presence of a few proteins at very high concentrations complicate the deep profiling of the human plasma proteome. Indeed, the concentrations of individual proteins span 10–12 orders of magnitude and abundant proteins, such as albumin, represent more than 99% of the total bulk mass of protein content [5,6]. In shotgun proteomics, the most abundant tryptic peptides dominate the MS analysis and hide the less abundant ones, which often derive from proteins of biological importance.

The removal of abundant proteins with immuno-affinity baits can significantly reduce the dynamic range and also the complexity of plasma protein samples [4,5,7,8]. Additional fractionation methods, either electrophoresis- or chromatography-based, are used to improve coverage of the plasma proteome. Nonetheless, all these sample preparation steps add to the overall analysis time, both at sample handling and subsequent mass spectrometric level. To analyze a large number of samples, sufficient throughput is necessary, limiting the applicability of extensive fractionation before MS analysis.

We evaluated several proteomic workflows for the analysis of human blood plasma samples. We used a commercial source of plasma obtained from pooled human blood. First, depletion of abundant proteins was assessed. One or two/tandem depletion stages (*i.e.*, immuno-affinity removal of some abundant proteins) were performed as previously reported by others [5,7]. For the so-called tandem depletion, the plasma sample was loaded a first time on the depletion cartridge and the flow-through was loaded again on the same depletion cartridge after its regeneration. Second, the duration of the gradient and the length of the column used for on-line reversed-phase liquid chromatography (RP-LC) before MS were varied (*i.e.*, 70 min separation with a 15 cm column and 150 min separation with a 50 cm column). At last, two dimensional (2D) RP/RP-LC with first dimension fractionation of the sample at basic pH, followed by a second at the usual acidic pH, was evaluated for deeper proteome coverage [9]. To assess relative protein quantification between samples, both label-free and isobaric tagging approaches were considered. All these workflows were discussed in terms of efforts (*i.e.*, invested time and cost), throughput, and plasma proteome coverage as an outcome. We argue that advised combination of such technologies is relevant for large clinical studies in the context of biomarker discovery.

2. Materials and methods

2.1. Material

Iodoacetamide (IAA), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), triethylammonium hydrogen carbonate buffer (TEAB) 1 M pH = 8.5 and sodium dodecyl sulfate (SDS) were purchased from Sigma (St. Louis, MO, USA). Formic acid (FA, 99%) was from BDH (VWR International Ltd., Poole, UK). Hydroxylamine solution 50 wt% in H_2O (99.999%) was from Aldrich (Milwaukee, WI, USA). Water (18.2 MΩ cm at 25 °C) was obtained from a Milli-Q apparatus (Millipore, Billerica, MA, USA) and acetonitrile was from BDH. The PROT20S kit was purchased from Sigma and used for abundant-protein depletion. The sixplex tandem mass tags (TMTs) were purchased from Thermo Scientific (Rockford, IL, USA). Sequencing grade modified trypsin was from Promega (Madison, WI, USA). Lyophilized Human plasma (Sigma, Lot 070M7009) containing 3.8% trisodium citrate as anti-coagulant, and obtained from pooled human blood, was re-suspended in 1 mL H_2O. Aliquots were kept at −20 °C.

2.2. Sample preparation

Human plasma sample was depleted using the PROT20S kit according to the manufacturer's instructions with 10 consecutive depletions of 8 μL of plasma. The protein recovery before and after one or two stages of depletion appears in Table 1. The protein concentration in the studied samples was determined at 280 nm in triplicate with a NanoDrop apparatus (Thermo Scientific) [10]. Reduction, alkylation, and protein digestion was performed as previously reported [11]. Briefly, each sample (*i.e.*, 10 to 100 μg of lyophilized proteins) was dissolved in 95 μL of TEAB 100 mM and 5 μL of SDS 2%. A volume of 5.3 μL TCEP 20 mM was added and incubation was performed for 1 h at 55 °C. A volume of 5.5 μL IAA 150 mM was added (incubation for 1 h in the dark). Enzymatic digestion was performed by addition of 10 μL trypsin at 0.25 $\mu g\,\mu L^{-1}$ in TEAB 100 mM and incubation overnight at 37 °C.

TMT labelling was performed by addition of 0.8 mg sixplex TMT reagent in 41 μL CH_3CN (incubation for 1 h at room temperature). After reaction, a volume of 8 μL hydroxylamine 5% in H_2O was added to each tube to react for 15 min.

All samples were purified with Oasis HLB cartridges (1cc, 30 mg) from Waters (Milford, MA, USA) followed by strong cation-exchange solid-phase extraction using home-prepared columns packed with SP Sepharose Fast Flow (Sigma). Samples were then evaporated to dryness before storage at −20 °C.

2.3. RP-LC and RP/RP-LC MS/MS

The samples were dissolved in H_2O/CH_3CN/FA 96.9/3/0.1 for RP-LC MS/MS analysis. LC MS/MS was performed on an LTQ orbitrap (OT) Elite from Thermo Scientific (San Jose, CA, USA) equipped with an Easy-nLC 1000 (Thermo Scientific) for 70 min separation or an Ultimate 3000 RSLC nano system

Table 1 – Protein amounts and yields obtained after 1stage or 2 stages of abundant-protein depletion. Starting from 80 μL of crude plasma, 10 depletion processes of 8 μL crude plasma were performed with the ProteoPrep® plasma immunodepletion kit from Sigma (see Section 2) to obtain the pooled depleted plasma sample, so-called 95% depleted. A unique second depletion stage was performed on half of the plasma sample 95% depleted, to obtain the depleted plasma sample, so-called 99% depleted.

Sample	Volume (μL)	Concentration (μg μL^{-1})	Amount (μg)	Yield (%)
Plasma	80	43.72 ± 0.35	3497	100
Plasma, 95% depleted	620[b]	0.847 ± 0.02	525	15
Plasma, 99% depleted[a]	220[b]	0.40 ± 0.01	87	5

[a] The equivalent of half of the sample followed the second depletion process.
[b] Volumes obtained after buffer exchange. These values were obtained from NanoDrop readouts (see Section 2).

(Thermo Scientific) for 150 min separation. Proteolytic peptides (~1 μg) were trapped on an Acclaim Pepmap 75 μm × 2 cm (C18, 3 μm, 100 Å) pre-column (Thermo Scientific). Following washing, peptides were separated on an Acclaim PepMap RSLC 75 μm × 15 or 50 cm (C18, 2 μm, 100 Å) column (Thermo Scientific) coupled to a stainless steel nanobore emitter (40 mm, OD 1/32″) (Thermo Scientific). The analytical separation was run either for 70 or 150 min using a gradient that reach 30% of CH_3CN after 60 and 140 min respectively and 80% after 70 and 150 min respectively. A flow rate of 220 nL min^{-1} was used. For MS1 survey scans, the OT resolution was 60,000 and the ion population was 1×10^6 with an *m/z* window from 300 to 1500. For MS2 in the LTQ with collision induced dissociation (CID), the ion population was 1×10^4 (isolation width of 2 *m/z* units) with a maximum injection time of 150 ms. For MS2 detection in the OT with higher-energy collisional dissociation (HCD), the ion population was set to 1×10^5 (isolation width of 2 *m/z* units), with a resolution of 15,000, first mass at *m/z* = 100 Th, and a maximum injection time of 750 ms. A maximum of 20 precursor ions (most intense) were selected for CID activation. CID was performed at 30% of the normalized collision energy (NCE). A combined CID/HCD method with a maximum of 10 precursor ions (most intense) was used to analyze TMT-labelled samples as previously described [11,12]. In this case, CID was performed at 30% of the NCE while HCD was performed at 60%. Dynamic exclusion was set for 60 s within a ±5 ppm window. The lock mass at *m/z* = 445.1200 Th was used. Each sample was analyzed in triplicate.

For upstream sample fractionation of the sample with RP-LC at basic pH (pH = 9.6), ~24 μg of each plasma protein digest (previously diluted at 2.4 μg μL^{-1} in ammonium formate 20 mM, pH 9.6 (solvent C)) was loaded on a 300 μm × 15 cm Acclaim PA II column (C18, 3 μm, 120 Å) from Thermo Scientific. The separation was performed with an Ultimate 3000 RSLC NCP system (Thermo Scientific) for 38 min with a gradient of solvent C and solvent D (ammonium formate 20 mM, pH 9.6 with 80% CH_3CN) at a flow rate of 6 μL min^{-1}. The gradient was performed as follows: 0–2 min 95% C and 5% D, then to 82% C and 18% D at 10 min, 60% C and 40% D at 15 min, 10% C and 90% D from 18 to 25 min, and 95% C and 5% D from 25 to 38 min. Twelve fractions were collected from 12 to 27 min. Samples were then evaporated to dryness before storage at −20 °C. Each fraction was diluted in 20 μL of H_2O/CH_3CN/FA 96.9/3/0.1 (see above). One fourth (*i.e.*, 5 μL of each fraction) was analyzed by RP-LC MS/MS with a 70 min gradient in duplicate (see above for details).

2.4. *Data analysis*

Proteome Discoverer 1.3 (version 1.3.0.339, Thermo Scientific) was used as raw data post-processing interface with the possibility to select scan events for peptide/protein identification and/or quantification. Identification was performed in the Swiss-Prot database (SwissProt_2012_07) with Homo sapiens taxonomy (20,232 sequences). Mascot (version 2.4.0, Matrix Sciences, London) was used as search engine. Variable amino acid modification was oxidized methionine. Carbamidomethylation of cysteines was set as fixed modification. In case of TMT sixplex labelled samples, TMT-labelled peptide amino terminus and TMT-labelled lysine (+229.163 Da) were set as fixed modifications. Trypsin was selected as the enzyme, with one potential missed cleavage. Peptide and fragment ion tolerance was respectively 10 ppm and 0.6 Da. All files were loaded in Scaffold 3.6.5 (Proteome Software, Portland, OR, USA) to visualize and validate the results. Peptide and protein thresholds were 95.0% and 99.0% respectively with a 2 unique peptide criterion to report protein identification (see Supplementary data).

3. Results

In this study, 3 sample preparation workflows were conducted and compared using aliquots of the same commercial pooled human plasma sample: (i) one consisted of a direct processing of the sample without any abundant-protein depletion; (ii) the second included abundant-protein depletion before further processing of the plasma sample and LC MS/MS analysis; and, (iii) the third workflow comprised two consecutive immunodepletion steps (*i.e.*, successive load and reload of the sample on the same depletion cartridge; see Section 2) to increase the efficiency of abundant-protein removal. For each sample, proteins were then reduced, alkylated and digested with trypsin. Half of the samples was labelled with a sixplex tandem mass tag (TMT) reagent, while the other half was kept unmodified as in a label-free approach. As illustrated in Fig. 1, six different samples were obtained and purified as described in the Material and Methods section before further LC MS analysis.

3.1. *Abundant-protein immunodepletion*

A commercial kit from Sigma was used. This kit contained a spin column system filled with affinity-purified polyclonal IgGs and small single-chain antibody ligands attached

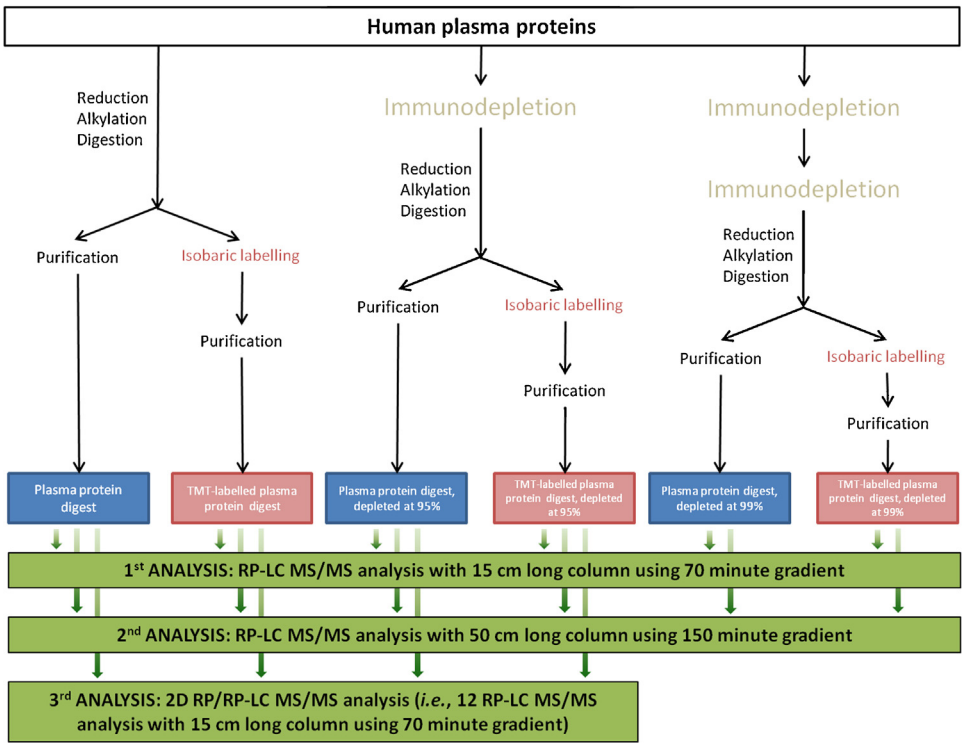

Fig. 1 – General workflows followed for the proteomic analyses of human plasma samples.

to agarose and allowed the depletion of 20 highly abundant proteins from human plasma or serum. The amount of proteins was measured before and after each stage of immunodepletion with a NanoDrop apparatus. According to the manufacturer, one stage of immunodepletion is able to remove 95% of the 20 highly abundant proteins while two stages of depletion remove 99% of these proteins. Using our commercial plasma sample, we found that 85% and 95% of the total bulk mass of protein content were removed after one or two consecutive depletion steps, respectively (Table 1). The concentration determined afterwards on the crude plasma sample with a Bradford assay [13] was found to be 14% lower (data not shown). The efficiency of the immunodepletion process was evaluated with LC MS analysis.

3.2. *RP-LC and RP/RP-LC MS/MS*

Different LC MS/MS analysis methodologies were applied to the samples (Fig. 1): 1D-RP-LC MS/MS with a 70 min gradient; 1D-RP-LC MS/MS with a 150 min gradient; and 2D-RP/RP-LC MS/MS resulting in an analysis time of about 20 h in total. The 2D-RP/RP-LC MS/MS workflow was not applied to the doubly depleted samples because of the low amounts of sample available after the two depletion stages.

Non-labelled samples were analyzed using a hybrid linear ion trap-orbitrap (LTQ-OT) mass spectrometer with a data-dependent MS/MS acquisition method that targeted the 20 most intense *m/z* signals in the MS survey scans to be further fragmented with collision-induced dissociation (CID). TMT-labelled samples were analyzed with a method that targeted the 10 most intense peaks in the MS survey scans to be further fragmented with CID and higher-energy collisional dissociation (HCD) as reported before for the relative protein quantification with isobaric tags [12].

Every workflow was evaluated with respect to the total number of protein identifications and unique peptide matches. The results of the 1D-RP-LC MS/MS with a 70 min gradient are given in Table 2a. The best proteome coverage was obtained when depletion was performed. The same observation was true for unique peptide matches. Indeed, for the label-free approach, 1532, 2166, and 1974 unique peptides were obtained on average for the non-depleted, singly depleted (so-called 95% depleted according to manufacturer), and doubly depleted (so-called 99% depleted) samples respectively. For the TMT-labelled samples, 845, 1221, and 1097 unique peptides were obtained on average for the non-depleted, 95% depleted, and 99% depleted samples, respectively.

The results of the 1D-RP-LC MS/MS with a 150 min gradient are given in Table 2b. The same tendencies as for the 1D-RP-LC MS/MS with a 70 min gradient were observed, namely that a higher number of proteins was obtained with depletion and that the label-free approach provided better proteome coverage with respect to the isobaric labelling technology using the LTQ-OT instrument. For the label-free approach, 2087, 2773, and 2468 unique peptides were obtained for the non-depleted, 95% depleted, and 99% depleted samples respectively. For the TMT-labelled samples, 1292, 1778, and 1580 unique peptides were obtained on average for the non-depleted, 95% depleted, and 99% depleted samples respectively.

The overlaps between the approaches in terms of proteome coverage and, in particular, the impact of the depletion of abundant plasma proteins were evaluated. The Venn diagrams of Fig. 2 present the number of identified proteins and their overlaps for both the label-free and TMT-based approaches.

Table 2a – Analysis of plasma samples using a 70 min gradient RP-LC MS/MS workflow (results obtained from 3 LC–MS/MS replicates).

Sample	Sample volume/μL[a]	Protocol description	Sample preparation time/hours[b]	Quantitative approach	Multiplexing	LC-MS/MS analysis time/min[c]	Number of identified proteins	Number of unique peptides
Plasma	2.3	Digestion with trypsin, purification	23	Label-free	No	100	120 ± 4%	1532 ± 1%
Plasma, 95% depleted	15.3	Depletion, digestion with trypsin, purification	30				154 ± 1%	2166 ± 1%
Plasma, 99% depleted	46	Depletion (2×), digestion with trypsin, purification	32				151 ± 2%	1974 ± 1%
Plasma, TMT-labelled	2.3	Digestion with trypsin, labelling, purification	24	Isobaric labelling	Yes		80 ± 6%	845 ± 17%
Plasma, 95% depleted, TMT-labelled	15.3	Depletion, digestion with trypsin, labelling, purification	32				102 ± 3%	1221 ± 5%
Plasma, 99% depleted, TMT-labelled	46	Depletion (2×), digestion with trypsin, labelling, purification	34				98 ± 6%	1097 ± 8%

[a] Sample volume to have 100 μg proteins available for protein digestion (calculated from Table 1).
[b] Including bench work, incubations, overnight digestion, and evaporation processes.
[c] Total time including sample loading/washing, analytical separation and re-equilibration of the LC columns.

Table 2b – Analysis of plasma samples using a 150 min gradient RP-LC MS/MS workflow (results obtained from 3 LC–MS/MS replicates).

Sample	Sample volume/μL[a]	Protocol Description	Sample preparation time/hours[b]	Quantitative approach	Multiplexing	LC-MS/MS analysis time/min[c]	Number of identified proteins	Number of unique peptides
Plasma	2.3	Digestion with trypsin, purification	23	Label-free	No	180	154 ± 2%	2087 ± 4%
Plasma, 95% depleted	15.3	Depletion, digestion with trypsin, purification	30				192 ± 1%	2773 ± 3%
Plasma, 99% depleted	46	Depletion (2×), digestion with trypsin, purification	32				194 ± 1%	2468 ± 1%
Plasma, TMT-labelled	2.3	Digestion with trypsin, labelling, purification	24	Isobaric labelling	Yes		110 ± 3%	1292 ± 1%
Plasma, 95% depleted, TMT-labelled	15.3	Depletion, digestion with trypsin, labelling, purification	32				129 ± 0%	1778 ± 1%
Plasma, 99% depleted, TMT-labelled	46	Depletion (2×), digestion with trypsin, labelling, purification	34				123 ± 2%	1580 ± 2%

[a] Sample volume to have 100 μg proteins available for protein digestion (calculated from Table 1).
[b] Including bench work, incubations, overnight digestion, and evaporation processes.
[c] Total time including sample loading/washing, analytical separation and re-equilibration of the LC columns.

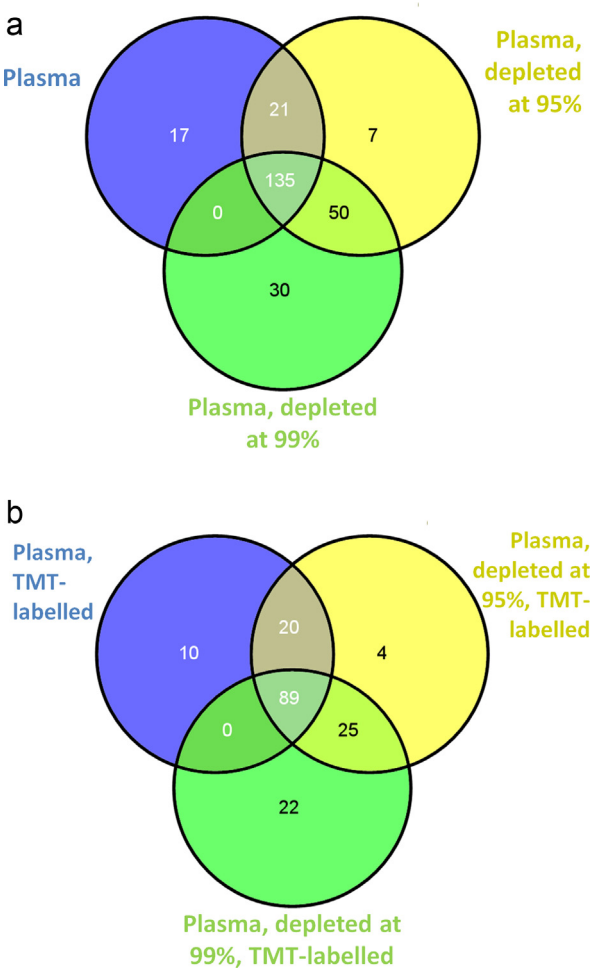

Fig. 2 – Number of proteins identified using RP-LC MS/MS with 150 min gradient RP-LC MS/MS workflow. Both label-free (a) and TMT-based (b) approaches were assessed. Plasma protein samples were prepared without any abundant-protein depletion (Plasma), or with abundant-protein depletion performed once (Plasma, 95% depleted) or twice (Plasma, 99% depleted).

Overall, 135 and 89 proteins were found in common between the non-depleted, 95% depleted, and 99% depleted samples for the label-free and TMT-based approaches respectively, corresponding to an overlap of 52% and 51% with respect to all identified proteins. In general, the tandem depletion (*i.e.*, 99% depleted samples) provided the highest numbers of new/unique protein identifications with 30 and 22 proteins exclusively identified with the label-free and TMT-based approaches respectively. Interestingly, there was no protein identification uniquely common to the non-depleted and 99% depleted samples.

Finally, a 2D-RP/RP-LC MS/MS workflow was applied. Because of the low sample amount left after the double depletion (*i.e.*, 87 μg; see Table 1), the doubly depleted plasma sample (*i.e.*, 99%) was not analyzed with 2D-RP/RP-LC MS/MS. 2D-RP/RP-LC MS/MS enabled a deeper coverage of the plasma proteome and a total 288 and 392 proteins were identified in the non-labelled depleted and 95% depleted plasma samples

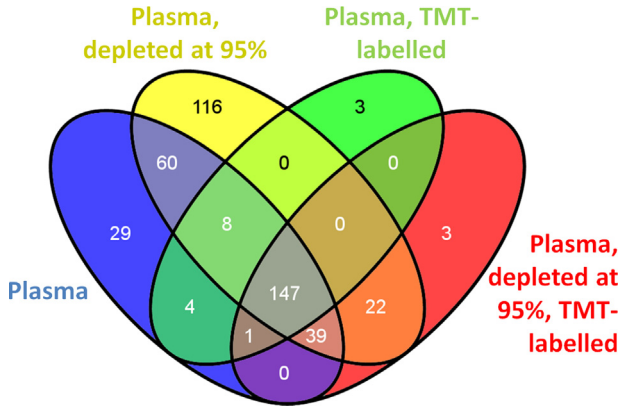

Fig. 3 – Number of proteins identified using a 2D-RP/RP-LC MS/MS workflow. Both label-free and TMT-based approaches were assessed and compared. Plasma protein samples were prepared without any abundant-protein depletion (Plasma and Plasma, TMT-labelled) or with abundant-protein depletion performed a single time (Plasma, 95% depleted and Plasma, 95% depleted, TMT-labelled).

within duplicate MS analyses (Fig. 3, Table 2c and Supplementary data). The Venn diagram of Fig. 3 compares the results of the 2D-RP/RP-LC MS/MS analyses. An overlap of 147 proteins was found between all these analyses.

Altogether, these proteomic analyses identified 437 protein entries with stringent criteria in the studied human plasma sample (see Supplementary data). Good sequence coverage was achieved with an average of more than 12 unique peptide matches per protein and a median of 7 unique peptides per protein. Classical plasma proteins were identified. We compared our protein list with the one recently reported by Farrah et al. [14].

4. Discussion

This work aimed at providing guidance for the choice of proteomics workflows to analyze human plasma samples and several workflows were assessed. Single immunodepletion combined with the use of longer RP-LC gradients and columns demonstrated proteome coverage of 192 and 129 proteins on average for the label-free and label-based approaches, respectively, at a manageable effort and throughput (*i.e.*, RP-LC MS/MS analysis of 3 h per sample) when applied to hundred or more samples, *e.g.*, in clinical studies.

4.1. *Investment vs. total protein identifications*

2D-RP/RP-LC MS/MS was shown to result in the best proteome coverage (*i.e.*, number of protein identifications). RP/RP-LC was previously shown to be a powerful alternative to the classical strong-cation exchange (SCX)/RP-LC [9]. In particular, the RP/RP-LC workflow eliminates the need for sample desalting, resulting in reduced sample losses and processing time together with a better resolution and peak capacity than with SCX [23]. Nonetheless, the long analysis time was a

Table 2c – Analysis of plasma samples using a 2D-RP/RP-LC MS/MS workflow (results obtained from 2 LC–MS/MS replicates).

Sample ID	Sample volume/μL[a]	Protocol Description	Sample preparation time/hours[b]	Quantitative approach	Multiplexing	LC-MS/MS analysis time/min[c]	Number of identified proteins	Number of unique peptides[e]
Plasma	2.3	Digestion with trypsin, purification, LC fractionation	25	Label-free	No	1200	256 ± 9%	3335 ± 5%
Plasma, 95% depleted	15.3	Depletion, digestion with trypsin, purification, LC fractionation	32				349 ± 6%	4163 ± 3%
Plasma, TMT-labelled	2.3	Digestion with trypsin, labelling, purification, LC fractionation	26	Isobaric labelling	Yes		150 ± 3%	1986 ± 5%
Plasma, 95% depleted, TMT-labelled	15.3	Depletion, digestion with trypsin, labelling, purification, LC fractionation	34				189 ± 10%	2301 ± 1%

[a] Sample volume to have 100 μg proteins available for protein digestion (calculated from Table 1).
[b] Including bench work, incubations, overnight digestion, and evaporation processes.
[c] Total time including sample loading/washing, analytical separation and re-equilibration of the LC columns to analyze the 12 fractions.

significant drawback that precludes the general application of such upstream fractionation in large clinical studies.

The better proteome coverage obtained with the label-free approach may depend on the speed of the MS/MS acquisition method (see Section 2). This fact is illustrated by the number of CID tandem mass spectra performed during analysis. For example, the average numbers of CID scans/spectra that were performed and used for peptide matching by the database search engine were 40′569 and 29′840 for the non-labelled and TMT-labelled plasma samples respectively (analysis with 1D-RP-LC MS/MS using a 150 min gradient). On the other hand, the quantification with TMTs offers multiplexing that significantly reduces the overall analysis time and, because the labelling occurs early in the analytical workflow the experimental variance for the quantification is reduced [15]. These advantages are highly relevant in the context of large clinical discovery studies. In terms of investment (which include instrument time/depreciation, and reagent costs), we found that – despite the investment in the tagging reagents – the sixplex isobaric tagging approach was about 10% less expensive than the label-free strategy. The decrease of the analysis time is indeed roughly inversely proportional to the multiplexing capabilities of the isobaric tags. On the top of that, if consider the workforce needed to analyze the samples and process the data, the cost savings should be even larger with the multiplexed approach.

The immunodepletion and RP-LC MS/MS analysis are considered as the limiting steps in terms of throughput while the rest of the sample preparation can be performed in a parallelized fashion (*i.e.*, reduction/alkylation/digestion, labelling, and purification). Immunodepletion can be performed with the column mounted onto an LC system but – as with RP-LC MS/MS analysis – only one sample can be processed at a time. Multiplying the number of instruments (LC systems and mass spectrometers) remains a very costly option while the possibility to label at protein level and enhanced multiplexing capabilities may be cost-efficient alternatives [16–18].

With regard to the reproducibility of the results, we assessed here the LC MS/MS experiments and showed that the numbers of both identified proteins and unique peptides varied less than 5% on average between technical replicates. The reproducibility, the pros and the cons of the individual analytical techniques (*i.e.*, immunodepletion, tryptic digestion, isobaric tagging, and LC fractionation) were not evaluated herein but were reported and discussed previously [9,19–23]. The use of internal standards is important in that respect to not only evaluate but also correct for experimental bias and to normalize the data.

4.2. *Investment vs. specific protein identifications*

As expected, some proteins were only identified with one or another of the depletion strategies (*i.e.*, absence of depletion, 95% depletion, and 99% depletion) (Fig. 2a). The 17 proteins identified uniquely in the non-depleted samples were mainly immunoglobulin chains in addition of polymeric immunoglobulin receptor (PIGR), titin (TITIN) and complement C1q subcomponent subunit A (C1QA) for the label-free approach. Likewise, for the tagging-based strategy, 7 immunoglobulin chains were identified in the non-depleted

samples only, in addition to complement C1q subcomponent subunit B (C1QB), dynein heavy chain 10 (DYH10), and transthyretin (TTHY) (Fig. 2b). These results appeared consistent as the immuno-based depletion targeted the removal of albumin, IgG, TTHY, fibrinogen, IgA, α2-marcroglobulin, IgM, α1-antitrypsin, complement C3, haptoglobulin, apolipoprotein A1, apolipoprotein A3, apolipoprotein B, α1-acid glycoprotein, ceruloplasmin, CO4B, C1QA, IgD, prealbumin, and plasminogen. But it clearly showed that immunoaffinity depletion efficiency is partial only. The identification of DYH10 was rather unexpected as it was not targeted by the immunoaffinity capture. This observation probably resulted from the stochastic nature of the data-dependant MS acquisition methodology. Despite the fact that the removal of the 20 proteins listed above appeared to be efficient, the concomitant unspecific removal of other proteins cannot be entirely excluded but was not evaluated in the present work [24]. The reader is referred to the manufacturer's specifications. Many proteins were uniquely identified when depletion was performed (see Supplementary data for the complete list). For instance, intercellular adhesion molecule 1 (ICAM1) was identified. ICAM1 was previously reported in the ng mL^{-1} range in human plasma samples as molecular markers for atherosclerosis [25] and stroke [26,27]. Peroxiredoxin-2 (PRDX2), which was recently reported to be increased in the microdialysates of the infarct core of stroke patients, was another example (Fig. 2a) [28] as well as prostaglandin-H2 D-isomerase (PTGDS) which was shown previously to be linked to brain disorders in cerebrospinal fluid [29].

4.3. *Proteome coverage overlap between workflows*

The protein overlap between label-free and isobaric labelling approaches was relatively small when using 2D-RP/RP-LC MS/MS (*i.e.*, 147 proteins). This observation was mainly due to the fact that the 2D-RP/RP-LC MS/MS analysis of the crude plasma labelled with TMT only identified 163 proteins (Fig. 3). With a higher degree of upstream fractionation, the label-free approach proved again superiority in terms of qualitative results (*i.e.*, number of peptides and protein identified). Nonetheless, the quantitative advantages offered by isobaric tagging when upstream fractionation is performed must be pointed out. The 212 proteins identified in the 95% depleted plasma sample labelled with TMT would have also been quantifiable straightforwardly.

Over the 437 protein entries found in this dataset, 353 have been recently reported in the human plasma PeptideAtlas reference set that encompassed 1929 canonical protein sequences [14]. Thirty one of these shared proteins between these datasets were reported in the ng mL^{-1} range in human plasma by Farrah et al. [14].

5. Conclusions

An evaluation of several proteomic workflows was performed for the analysis of human plasma samples. Extensive sample preparation and fractionation are necessary to significantly increase proteome coverage. However, investments must be limited when it comes to realistic workflow deployment in large clinical studies. Multi-affinity depletion of highly abundant proteins was efficient to access lower abundance proteins in plasma. Proteins at a concentration of ng mL^{-1} range were confidently identified. The use of longer RP-LC gradients and columns appeared as a very valuable solution in terms of a good compromise between effort, throughput and reproducibility. While this option increases LC MS/MS analytical time, it does not require any supplemental human workforce and sample handling. As regards the relative protein quantification between samples, isobaric tagging is recommended because of the multiplexing capabilities that significantly reduce the overall analysis time and technical variability. For large-scale clinical studies the following workflow appears to us both feasible and efficient: (i) abundant-protein immunodepletion; (ii) reduction/alkylation/digestion of the proteins; (iii) isobaric labelling of the peptides; and, (iv) RP-LC MS/MS analysis of the samples using long gradients and columns. We are now evaluating the analytical performance of a complete, integrated human plasma proteomics workflow (based on the elements described and assessed in this paper) and have automated most of the sample preparation steps using LC systems for the immunodepletion and liquid handling robotics for the reduction/alkylation/digestion, isobaric labelling, and sample purification.

Acknowledgements

We thank John Corthésy and Antonio Núñez Galindo for discussion and technical support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.euprot.2013.08.001.

REFERENCES

[1] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1:845–67.
[2] Anderson NL. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. Clin Chem 2010;56:177–85.
[3] Whiteaker JR, Lin C, Kennedy J, Hou L, Trute M, Sokal I, et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. Nat Biotechnol 2011;29:625–34.
[4] Whiteaker JR, Zhang H, Eng JK, Fang R, Piening BD, Feng LC, et al. Head-to-head comparison of serum fractionation techniques. J Proteome Res 2007;6:828–36.
[5] Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, et al. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. Mol Cell Proteomics 2008;7:1963–73.
[6] Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol 2009;3:33–44.
[7] Shuford CM, Hawkridge AM, Burnett JC, Muddiman DC. Utilizing spectral counting to quantitatively characterize tandem removal of abundant proteins (TRAP) in human plasma. Anal Chem 2010;82:10179–85.

[8] Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res 2010;9:4982–91.
[9] Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci 2005;28:1694–703.
[10] Desjardins P, Hansen JB, Allen M. Microvolume spectrophotometric and fluorometric determination of protein concentration. Curr Protein Pept Sci 2009, 3.10.1–3.10.6.
[11] Dayon L, Sonderegger B, Kussmann M. Combination of gas-phase fractionation and MS 3 acquisition modes for relative protein quantification with isobaric tagging. J Proteome Res 2012;11:5081–9.
[12] Dayon L, Pasquarello C, Hoogland C, Sanchez JC, Scherl A. Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. J Proteom 2010;73:769–77.
[13] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248–54.
[14] Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. Mol Cell Proteomics 2011;10:1.
[15] Christoforou AL, Lilley KS. Isobaric tagging approaches in quantitative proteomics: the ups and downs. Anal Bioanal Chem 2012;404:1029–37.
[16] Engmann O, Campbell J, Ward M, Giese KP, Thompson AJ. Comparison of a protein-level and peptide-level labeling strategy for quantitative proteomics of synaptosomes using isobaric tags. J Proteome Res 2010;9: 2725–33.
[17] McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem 2012;84:7469–78.
[18] Werner T, Becher I, Sweetman G, Doce C, Savitski MM, Bantscheff M. High-resolution enabled TMT 8-plexing. Anal Chem 2012;84:7188–94.
[19] Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 2008;80:2921–31.
[20] Hagman C, Ramström M, Jansson M, James P, Håkansson P, Bergquist J. Reproducibility of tryptic digestion investigated by quantitative fourier transform ion cyclotron resonance mass spectrometry. J Proteome Res 2005;4:394–9.
[21] Pernemalm M, Orre LM, Lengqvist J, Wikström P, Lewensohn R, Lehtiö J. Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. J Proteome Res 2008;7:2712–22.
[22] Smith MPW, Wood SL, Zougman A, Ho JTC, Peng J, Jackson D, 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 2011;11:2222–35.
[23] Yang F, Shen Y, Camp Ii DG, Smith RD. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev Proteom 2012;9:129–34.
[24] Roche S, Tiers L, Provansal M, Seveno M, Piva M-T, Jouin P, et al. Depletion of one, six, twelve or twenty major blood proteins before proteomic analysis: the more the better. J Proteom 2009;72:945–51.
[25] Hwang SJ, Ballantyne CM, Sharrett AR, Smith LC, Davis CE, Gotto Jr AM, et al. Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: The Atherosclerosis Risk In Communities (ARIC) study. Circulation 1997;96:4219–25.
[26] Kunz AB, Kraus J, Young P, Reuss R, Wipfler P, Oschmann P, et al. Biomarkers of inflammation and endothelial dysfunction in stroke with and without sleep apnea. Cerebrovasc Dis 2012;33:453–60.
[27] Volcik KA, Ballantyne CM, Hoogeveen R, Folsom AR, Boerwinkle E. Intercellular adhesion molecule-1 G241R polymorphism predicts risk of incident ischemic stroke: Atherosclerosis risk in communities study. Stroke 2010;41:1038–40.
[28] Dayon L, Turck N, García-Berrocoso T, Walter N, Burkhard PR, Vilalta A, et al. Brain extracellular fluid protein changes in acute stroke patients. J Proteome Res 2011;10:1043–51.
[29] Lescuyer P, Gandini A, Burkhard PR, Hochstrasser DF, Sanchez JC. Prostaglandin D2 synthase and its post-translational modifications in neurological disorders. Electrophoresis 2005;26:4563–70.