Amino Acids (2012) 43:2199–2202
DOI 10.1007/s00726-012-1267-1

SHORT COMMUNICATION

High confidence and sensitivity four-dimensional fractionation for human plasma proteome analysis

Renato Millioni · Serena Tolin · Gian Paolo Fadini · Marco Falda · Bas van Breukelen · Paolo Tessari · Giorgio Arrigoni

Received: 22 November 2011 / Accepted: 5 March 2012 / Published online: 21 March 2012

Abstract Reducing the complexity of plasma proteome through complex multidimensional fractionation protocols is critical for the detection of low abundance proteins that have the potential to be the most specific disease biomarkers. Therefore, we examined a four dimension profiling method, which includes low abundance protein enrichment, tryptic digestion and peptide fractionation by IEF, SCX and RP-LC. The application of peptide pI filtering as an additional criterion for the validation of the identifications allows to minimize the false discovery rate and to optimize the best settings of the protein identification database search engine. This sequential approach allows for the identification of low abundance proteins, such as angiogenin (10^{-9} g/L), pigment epithelium growth factor (10^{-8} g/L), hepatocyte growth factor activator (10^{-7} g/L) and thrombospondin-1 (10^{-6} g/L), having concentrations similar to those of many other growth factors and cytokines involved in disease pathophysiology.

Keywords Peptide isoelectrofocusing · Peptide pI filtering · False discovery rate · Plasma proteome

Electronic supplementary material The online version of this article (doi:10.1007/s00726-012-1267-1) contains supplementary material, which is available to authorized users.

R. Millioni (✉) · S. Tolin · G. P. Fadini · P. Tessari
Department of Medicine, University of Padua,
Padua, Italy
e-mail: millionirenato@gmail.com

R. Millioni · G. Arrigoni
Proteomics Centre of Padua University,
VIMM and Padua University Hospital, Padua, Italy

S. Tolin · G. P. Fadini · G. Arrigoni
VIMM, Venetian Institute of Molecular Medicine,
Padua, Italy

M. Falda · G. Arrigoni
Department of Biological Chemistry,
University of Padua, Padua, Italy

B. van Breukelen
Biomolecular Mass Spectrometry and Proteomics Group,
Bijvoet Centre for Biomolecular Research and Utrecht
Institute for Pharmaceutical Sciences, Utrecht University
and Netherlands Proteomics Centre, Padualaan 8,
3584 CH Utrecht, The Netherlands

The human plasma section of the Peptide Atlas database (http://www.peptideatlas.org/hupo/hppp) contains 1,929 identified proteins. Since this database collects the joint efforts of many different laboratories, the relatively low number of identifications well reflects the complexity of this proteome. Though a general consensus on a specific procedure to investigate the "hidden" plasma proteome does not yet exist, it is well established that better results can be obtained combining various separation techniques to reduce sample complexity (Hoffman et al. 2007). Here, we investigated a four-dimensional fractionation protocol for the analysis of plasma proteome. The experimental workflow is presented in Fig. 1.

The most used plasma pre-treatments for biomarker discovery are the immuno-subtraction of high abundance proteins and the enrichment of low abundance proteins using combinatorial peptide ligand libraries (CPLL). Recently, we showed that the depletion of the 20 most abundant proteins and the CPLL enrichment allowed the identification of proteins belonging to the same order of magnitude in terms of their plasma concentration (Millioni et al. 2011). However, the enrichment approach has the great advantage of obtaining much larger amount of

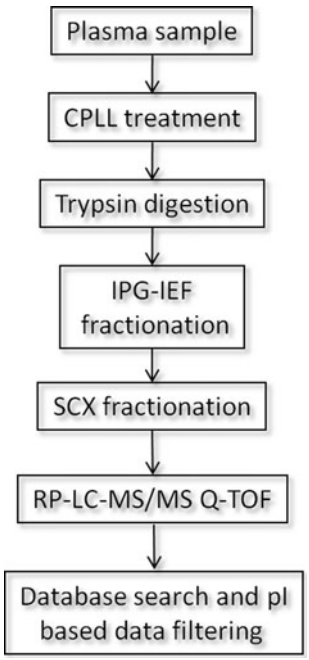

Fig. 1 Experimental workflow

material that can be further fractionated (Millioni et al. 2011). Furthermore, a detailed report of the improvements in proteomic metrics using CPLL prior to multidimensional protein identification technology (MudPIT) has been published more recently (Fonslow et al. 2011). Hence, in this study, we used the CPLL as the front-line step. Enriched proteins had then to be extensively fractionated to increase analysis sensitivity. The MudPIT, combining strong cation exchange (SCX) chromatography and reverse phase (RP) chromatography, is the most used shotgun approach for this purpose (Yates et al. 2009). However, new methods have recently emerged with the idea to modify the classical MudPIT approach using IPG-IEF in place of SCX (Cargile et al. 2005). It was demonstrated that peptide IEF had many advantages over SCX, such as greater loading capacity (Essader et al. 2005), reproducibility and resolution (Slebos et al. 2008). In addition, it allows using peptide pI values to identify peptide post-translational modifications (Lengqvist et al. 2011) and to reduce false positive identifications (Krijgsveld et al. 2006). Thanks to this reduction, the cross-correlation parameter (Xcorr) of SEQUEST can be lowered resulting in more identifications (Cargile et al. 2004). Furthermore, IPG-IEF peptide fractionation, just like SCX, is compatible with iTRAQ labeling for quantification (Lengqvist et al. 2007). Nevertheless, in silico studies on human proteome showed that even a narrow IPG fraction could hold up to several thousands of different peptides (Eriksson et al. 2008) that cannot be identified using only a RP chromatography step and a data-dependent acquisition mode (Michalski et al. 2011). Here, we investigated a protocol that combines IEF, SCX and RP to separate peptides on the basis of different properties, such as pI, net charge and hydrophobicity. To the best of our knowledge, this is the first report in which the fractionation protocol was examined for the analysis of plasma proteome. Detailed methods are reported in Supplementary Materials (Supplementary data 1). Briefly, CPLL enriched proteins were precipitated, reduced, alkylated and trypsin digested. The peptide sample (150 μg) was subjected to IPG-IEF. After focusing, the IPG strip was cut in eight parts. Peptides eluted by each strip part were loaded onto a SCX cartridge and stepwise eluted with four concentrations of KCl. The 32 peptide fractions obtained by IEF and SCX fractionations were analyzed by RP-LC–MS/MS using a Q-TOF mass spectrometer.

Theoretical values of pI peptides identified by SEQUEST and belonging to the same IEF fraction were calculated in batch using an algorithm developed by Gauci et al. (2008). Medium (FDR < 0.05) and high (FDR < 0.01) confidence identifications were considered as correct if the calculated pI of the peptides corresponded to the experimental pI within a ± 1 pH unit interval. We took advantage of pI filtering application to evaluate the quality of identifications obtained with four different settings of SEQUEST. Starting with a stringent setting (Set 1), SEQUEST identified a number of peptides (named "total identifications") from each IEF fraction. By pI filtering, we classified the peptides with a theoretical pI within or outside the selected pH interval as true positives (TP) and false positives (FP), respectively. We gradually "relaxed" the minimum allowed Xcorr values of SEQUEST (without changing the maximum allowed FDR) as explained in Supplementary materials (Supplementary data 1) and, for each set, we calculated the TP and FP values on the basis of pI filtering. Additional identifications of peptides with a pI within the expected range of pH were obtained. We named these newly identified peptides as "NewP". This comparison was performed using data from fraction IEF-1 and 4 (Table 1), and the best values for Xcorr found using this method where then applied to analyze all fractions. Peptide pI filtering can help to assess the performance of different SEQUEST settings and consequently to select the best setting to use. With this strategy, it is possible to significantly increase the identifications, still maintaining the analysis accurate since most of the FP can be identified and eliminated. A limitation of this approach is that a number of erroneous peptides could have a pI within the expected pH range and therefore may be incorrectly included among the TP. For this reason, the SEQUEST filters cannot be arbitrarily lowered, but they have to be determined looking for a condition where the increase of new identifications exceeds that of FP. Another important parameter to evaluate the settings is the precision [TP/(TP + FP)], i.e., a

Table 1 Parameters for the comparison of results obtained in fractions IEF-1 and 4 using four different settings of SEQUEST

IEF	SET	Total identification	FP	TP = total identification−FP	NewP	NewP/FP	Precision (%)
1	1	624	18	606	N/D	N/D	97
	2	659	23	636	30	1.3	96
	3	693	31	662	56	1.8	95
	4	767	66	701	95	1.4	91
4	1	314	13	301	N/D	N/D	96
	2	348	16	332	31	1.94	95
	3	376	17	359	58	3.41	95
	4	433	33	400	99	3.00	92

measure of result fidelity. However, in biomarker discovery studies, the primary aim is to obtain as many identifications as possible and therefore a small reduction in precision can be tolerated. Based on these considerations, we decided to use the Set 3 since it was the one with the highest ratio between NewP and FP values and with a good, even if not optimal, precision (Table 1).

False positive peptides included both SEQUEST erroneous matched peptides and unfocused peptides. This can be argued by looking at the uneven distribution of FP among the IEF fractions (Fig. 2). The ratio between FP and TP was higher in the more basic fractions and this is probably due to the lower efficacy of immobilines in this pH region, an issue already reported by others (Krijgsveld et al. 2006; Hubner et al. 2008).

Analyzing all the fractions, the application of the pI filtering revealed that 583 out of 3,700 peptides returned by SEQUEST as positive hits had to be discarded. After removing these FP, the number of identified protein groups (characterized by at least 2 peptides) decreased from 503 to 417. This number is about five times higher than that reported in our previous study where the same protocol was used, but without the IPG-IEF fractionation step (Millioni et al. 2011). As we increased the number of fractions, a consequent

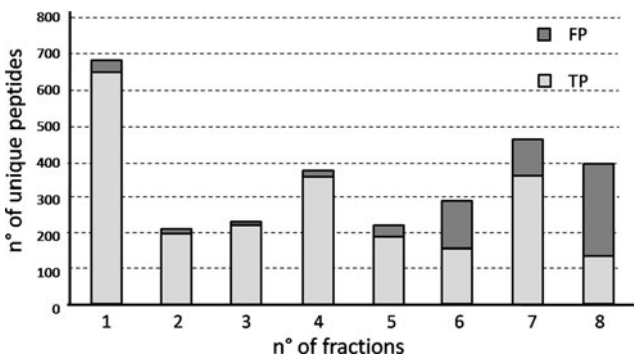

Fig. 2 *Bar graph* showing the number of unique peptides identified in each pH fraction. According to the pI filtering, *black* and *gray sections* represent, respectively, the true and the false positive identifications

increase of protein identifications was expectable. However, we can ascribe this result mainly to the combination of different orthogonal methods, since in a previous analysis (data not shown) the increase of SCX fractions, after CPLL enrichment of plasma, led to an increased protein coverage rather than to a higher number of identifications.

Validated and excluded proteins were further examined, to find possible correlation with other important parameters, such as the protein coverage and spectral count values. We found that the excluded protein groups had lower mean and median coverage, spectral count and number of peptides (Supplemental Table 1) with respect to the included protein groups. Some validated proteins had coverage and spectral counts values similar to those of the excluded proteins, which means that these parameters are suggestive but not sufficient criteria for exclusion. These findings further underline the ability of the peptide pI filtering process to increase the stringency of protein identifications.

A drawback of this workflow is the time required for the computational data analysis. Assuming an increase in the number of fractions to be analyzed, it becomes indispensable to find a solution to significantly shorten the data analysis time. The development of computer tools for batch processing of all the various phases of the analysis is mandatory to make the process easier and the data analysis faster.

When looking at the list of identified proteins (for details see "Supplementary data: protein and peptide identifications"), it appears that this sequential approach allows for the identification of low abundance proteins, such as angiogenin (10^{-9} g/L), pigment epithelium growth factor (10^{-8} g/L), hepatocyte growth factor activator (10^{-7} g/L) and thrombospondin-1 (10^{-6} g/L), having concentrations similar to those of many other growth factors and cytokines involved in disease pathophysiology.

Considering that the number of fractions analyzed in this study is relatively small and could be easily increased, the proposed workflow offers interesting possibilities for high sensitivity proteome profiling.

Conflict of interest The authors declare that they have no conflict of interest.

References

Cargile BJ, Bundy JL, Freeman TW, Stephenson JL Jr (2004) Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. J Proteome Res 3: 112–119

Cargile BJ, Sevinsky JR, Essader AS, Stephenson JL Jr, Bundy JL (2005) Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. J Biomol Tech 16:181–189

Eriksson H, Lengqvist J, Hedlund J, Uhlen K, Orre LM, Bjellqvist B, Persson B, Lehtio J, Jakobsson PJ (2008) Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms. Proteomics 8:3008–3018

Essader AS, Cargile BJ, Bundy JL, Stephenson JL Jr (2005) A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. Proteomics 5:24–34

Fonslow BR, Carvalho PC, Academia K, Freeby S, Xu T, Nakorchevsky A, Paulus A, Yates JR 3rd (2011) Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. J Proteome Res 10:3690–3700

Gauci S, van Breukelen B, Lemeer SM, Krijgsveld J, Heck AJ (2008) A versatile peptide pI calculator for phosphorylated and N-terminal acetylated peptides experimentally tested using peptide isoelectric focusing. Proteomics 8:4898–4906

Hoffman SA, Joo WA, Echan LA, Speicher DW (2007) Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. J Chromatogr B Analyt Technol Biomed Life Sci 849: 43–52

Hubner NC, Ren S, Mann M (2008) Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics 8:4862–4872

Krijgsveld J, Gauci S, Dormeyer W, Heck AJ (2006) In-gel isoelectric focusing of peptides as a tool for improved protein identification. J Proteome Res 5:1721–1730

Lengqvist J, Uhlen K, Lehtio J (2007) iTRAQ compatibility of peptide immobilized pH gradient isoelectric focusing. Proteomics 7:1746–1752

Lengqvist J, Eriksson H, Gry M, Uhlen K, Bjorklund C, Bjellqvist B, Jakobsson PJ, Lehtio J (2011) Observed peptide pI and retention time shifts as a result of post-translational modifications in multidimensional separations using narrow-range IPG-IEF. Amino Acids 40:697–711

Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC–MS/MS. J Proteome Res 10:1785–1793

Millioni R, Tolin S, Puricelli L, Sbrignadello S, Fadini GP, Tessari P, Arrigoni G (2011) High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. PLoS ONE 6:e19603

Slebos RJ, Brock JW, Winters NF, Stuart SR, Martinez MA, Li M, Chambers MC, Zimmerman LJ, Ham AJ, Tabb DL, Liebler DC (2008) Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography–tandem mass spectrometry. J Proteome Res 7:5286–5294

Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 11:49–79