Published in final edited form as:

Proteomics Clin Appl. 2012 June ; 6(5-6): 304–308. doi:10.1002/prca.201100107.

Precursor ion exclusion for enhanced identification of plasma biomarkers

Wells W. Wu¹, Rong-Fong Shen², Sung-Soo Park³, Bronwen Martin^{1,*}, and Stuart Maudsley^{3,*}

¹Metabolism Unit, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

²Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA

³Receptor Pharmacology Unit, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

Abstract

Purpose—Our study aims to establish a plasma biomarker analysis workflow, with fewer fractionation steps, for enhanced identification of plasma biomarkers by precursor ion exclusion (PIE).

Experimental design—Plasma samples were depleted for highly abundant proteins, then further fractionated by molecular weight (MW), before trypsinization for LTQ-Orbitrap mass analysis. Data-dependent acquisition (DDA) was used for baseline analysis. PIE involves the re-injection of samples with exclusion of the previously identified peptides. We compared analyses using multiple PIE iterations, compared to DDA, for plasma interrogation

Results—A higher percentage of unique plasma peptides was identified with PIE, compared to DDA. The first PIE iteration reveals an increase of 75–112% more peptides than the DDA method alone. PIE can interrogate complex plasma samples with the percentage of peptides identified successively increasing with even 4 iterations. The total number of peptides identified increases rapidly across the first three PIE iterations and then continues more slowly up to nine iterations.

Conclusions and clinical relevance—Iterative injections with PIE resulted in many more peptide identifications in plasma samples of varying degrees of complexity, compared to re-injections using similar DDA parameters. PIE methods may therefore expand our ability to recover plasma peptides for plasma biomarker discovery.

Keywords

Multi-dimensional fractionation; Plasma biomarkers; Precursor ion exclusion

Correspondence: Dr. Stuart Maudsley, Receptor Pharmacology Unit, National Institute on Aging, National Institutes of Health, 251 Bayview Blvd., Baltimore, MD 21224, USA, maudsleyst@grc.nia.nih.gov, **Fax:** +1-410-558-8323.

*These authors contributed equally to this work.

Colour Online: See the article online to view Fig. 3 in colour.

The authors have declared no conflict of interest.

1 Introduction

Human plasma is a highly complex fluid containing an extremely wide dynamic range of proteins (>10 orders of magnitude) [1]. In addition, approximately 99% of the plasma proteome content is comprised of 22 highly abundant proteins (HAPs) [2]. The presence of HAPs in plasma hinders mass spectrometric detection of potential biomarkers present at much lower levels (ng/mL) [3]. The removal of certain HAPs is often the first step in plasma biomarker studies [4,5]. However upon such removal, the vast dynamic range of plasma proteins that remain still renders biomarker identification a challenging task [5, 6]. Multi-dimensional fractionation could partially resolve this complex issue, but it reduces throughput rate and increases sample loss, as well as increasing inter-experimental variations. An intermediate level (two dimensions) of fractionation is more practical for the timely investigation of large-scale clinical sample sets [7].

Besides using fractionation, a mass spectrometric strategy to improve biomarker identification can be achieved by repeated sample analyses utilizing identical parameters in data-dependent acquisition (DDA). Depending on the sample complexity, higher abundant peptides may repeatedly trigger tandem mass spectrometry (MS2) in such a DDA approach, thus hindering the detection of lower abundant biomarkers. As a refinement to the DDA method, precursor ion exclusion (PIE) has been proposed. Essentially, PIE involves the reinjection of samples for re-analysis by LC-MS/MS with exclusion of the peptides identified previously. PIE has previously been tested for more comprehensive protein identification in samples less complex than plasma [8–10]. However, the effects of PIE on mapping the highly complex plasma proteome have not been comprehensively investigated. In this study, we systematically assessed the ability of PIE, with up to nine iterations, to interrogate plasma samples with an intermediate level of fractionation.

2 Materials and methods

Plasma samples ($n = 4$, mean age = 71.9 ± 0.79 years, sex = male) were obtained from the Baltimore Longitudinal Study of Aging (BLSA) [11,12] and pooled into one stock. For BLSA samples, the written approval of Institutional Review Board was obtained from the National Institute on Aging, and informed written consent was obtained from all participants. All collected samples were analyzed in a de-identified manner. HAPs were removed using an Agilent (Wilmington, DE, USA) human 14 multiple affinity removal spin (MARS) cartridge (thus denoted as “depleted plasma”) in the presence of $1\times$ Halt protease/phosphatase inhibitor cocktail (Pierce, Rockford, IL, USA). For selected experiments, aliquots of depleted plasma were deglycosylated with PNGase F (New England Biolabs, Ipswich, MA, USA) under a denaturing condition, according to the manufacturer’s instructions. Detergents (1% NP-40 and 0.5% SDS) were removed using Pierce detergent removal spin columns (ThermoFisher Scientific, Rockford, IL, USA). HCT116 cells, a human colon adenocarcinoma cell line obtained from American Type Culture Collection, were cultured in McCoy’s 5A media, and served as a control protein sample, known to possess less endogenous proteolytic cleavage than plasma samples, as previously described [13]. Nondepleted and depleted plasma samples, as well as HCT116 lysates were fractionated into 12 molecular weight (MW) fractions (Gelfree 8100, Protein Discovery, Inc.

Knoxville, TN, USA). Gelfree[®] (Expedeon/Protein Discovery, San Diego, CA, USA) is a novel apparatus based on MW fractionation in liquid phase to reduce complexity and offers a means to assess the level of proteolytic degradation in plasma samples. Proteins in each MW fraction were reduced, alkylated, and digested with trypsin [13]. Protein concentrations were determined using micro-bicinchoninic acid (BSA) assay kit (Pierce). Samples were analyzed by LC-MS/MS using an Eksigent NanoLC Ultra 2D (Dublin, CA, USA) and Thermo Fisher Scientific LTQ Orbitrap XL (San Jose, CA, USA). Briefly, ~1 µg peptides were first loaded onto a trap cartridge (Agilent Zorbax 300SB-C18, 5 µm, 300 µm id × 5 mm at 5 µL/mL), then eluted onto a reversed phase PicoFrit column (New Objective, Woburn, MA, USA Betabasic C18, 5 µm, 150 Å, 360 µm id × 75 µm id × 10 cm) using a linear 60-min gradient of ACN (2–62%) containing 0.1% formic acid at 250 nL/min flowrate. The eluted peptides were sprayed into the LTQ Orbitrap XL. The DDA mode was enabled, and each FTMS MS1 (Thermo Fisher Scientific) scan (60 000 resolution) was followed by four LTQ MS2 scans. The spray voltage and the ion transfer tube temperature were set at 1.8 kV and 180°C, respectively. Proteome Discoverer 1.1 (Thermo Fisher Scientific) was used to identify proteins at precursor and fragment ion tolerances of 25 ppm and 0.8 Da, respectively, at false discovery rate of 5%, and at least two peptides needed to be identified using the human RefSeq database. A conversion of N → D (deamidation) was set as a variable modification by Proteome Discoverer for identification of PNGase F deglycosylated peptides. PIE lists containing respective *m/z* and retention times were generated using Proteome Discoverer using the steps outlined in the Supporting Information. Plasma samples were interrogated using PIE with up to nine iterations.

3 Results and discussion

3.1 HAP removal and fractionation based on MW

Analysis of the box-(25th and 75th percentiles)-and-whisker (10th and 90th percentiles) distributions of proteins corresponding to different MW fractions for depleted plasma and HCT116 cell lysate (data in Supporting Information Tables S1 and S2, respectively) indicated a greater range variation for the depleted plasma compared to the expectedly more homogenous HCT116 cell lysates. The longer top-whiskers suggest a higher level of proteolytic degradation of proteins (i.e. proteins eluted earlier than expected) in plasma compared to the control of HCT116 lysate. The greater apparent proteolytic cleavage might also explain the wider box size of the plasma sample MW ranges (boxes) compared to the HCT116 lysates. Human plasma contains many protein fragments that are a result of endogenous proteolysis [14,15]. It has been reported that 55% of the circulating peptides identified in human blood are the fragments from higher MW plasma proteins [14]. The endogenous proteolytic cleavage often uniformly occurs in samples of a series of batches due to similar collection conditions. Sets of precursor ions of common proteolytic fragments in given batches could be determined as features in prior LC-MS/MS runs. PIE can use the exclusion mass lists to guide the mass spectrometer to avoid sequencing the endogenous degradants. The endogenous proteolysis might also cleave epitopes recognizable by the immobilized antibodies on MARS cartridges [6, 16] and explains why certain HAPs could be identified in the depleted plasma (Supporting Information Table S1) even in the presence

of protease/phosphatase inhibitors (Section 2). In addition to HAPs, the endogenous proteolysis could have caused other proteins to be eluted earlier than expected (Fig. 1A).

In addition, the plasma sample MW distribution possessed longer bottom-whiskers than the HCT116 cell lysates (Fig. 1A and B). We speculate that the bottom-whiskers (related to proteins eluted later than expected) are largely due to post-translational glycosylation. For example, 82 of the 100 most abundant plasma proteins are reported to commonly exist as functional glycoproteins [17]. Supporting Information Fig. S1 compares the box-and-whisker distributions of plasma proteins versus deglycosylated plasma proteins (data in Supporting Information Table S3). There are generally shorter distances between medians and bottom whisker 10th quartiles in the deglycosylated plasma than the non-deglycosylated counterpart.

Although there are other means of fractionation at the protein level, the graphic nature of Fig. 1 and Supporting Information Fig. S1 highlights the distinctive advantage of fractionation based on MW and the importance of including post-translational modifications (PTMs) in biomarker discovery. With peptide-based bottom-up approaches combined with the MW information, plasma biomarkers (intact or PTM proteins, and their proteolytic fragments) among different diseases or pathophysiological states could therefore be more accurately characterized and differentiated.

3.2 Comparison of PIE and DDA analysis of plasma samples of differing complexity

We first compared PIE and DDA methodologies using samples with differing complexity, i.e. non-depleted plasma, depleted plasma, and MW fractions 2–12 of the depleted plasma. An initial DDA run was acquired as the baseline, followed by PIE versus another identical DDA run, respectively. All of the samples demonstrated a higher percentage of unique peptides or proteins gained (relative to baseline) with PIE compared to DDA (Fig. 2, data in Supporting Information Table S4), with the only exception being the percentage of proteins obtained from the non-depleted plasma. We further investigated this specific anomalous result with the non-depleted plasma to test whether this lack of superiority of PIE versus DDA was due to the extreme complexity (no. of peptides eluted per unit of time, multiple peptides sequenced belonging to the same abundant protein, etc.) of this sample. For peptide identification, the first PIE for non-depleted, depleted, and MW fractions of the depleted plasma demonstrated an increase of 20.7, 111.6, and 74.5% (average), respectively, of more peptides identified than the DDA alone. Thus, we concluded that overall for proteomic samples, a greater degree of efficient data extraction was effected with PIE compared to DDA analyses.

Having demonstrated that the PIE analysis was more effective than DDA to probe unique peptides, we systematically assessed the effects upon resultant data streams generated using PIE with 4 iterations. Data represented in Fig. 3 (data in Supporting Information Table S4) demonstrate that PIE can interrogate all the plasma sample types studied, with the percentage of unique accumulated peptides identified showing no obvious plateau, even with 4 iterations (in some cases, up to nine iterations). The increasing peptide identifications, with further successive PIE iterations, were most pronounced for the non-depleted or depleted plasma samples compared to the various MW fractions. The total number of

identified peptides grows rapidly across the first three PIE runs and then begins to plateau with increasing PIE iterations up to nine. Supporting Information Fig. S2 (data in Supporting Information Table S4) summarizes the numbers of uniquely identified proteins from successively multiple PIE injections. New proteins are identified preferentially from the second or third injections (PIE1 and 2). It has been reported that successively less-intense precursor ions are sequenced over the course of multiple PIE runs; whereas the precursor ion intensity remains relatively constant using standard DDA [10]. Our data represented in Supporting Information Fig. S3 corroborate this concept.

4 Concluding remarks

In conclusion, the iterative injections with PIE resulted in the identification of considerably more peptides from plasma samples of varying degrees of complexity, compared to re-injections using identical parameters using standard DDAs. PIE methods are software-based and are easier and more cost effective to set up than multi-dimensional fractionations that can also incur deleterious sample loss. PIE, therefore, can access an expanded range of peptides (using plasma samples of different degrees of fractionation), thus facilitating a highly adaptable and efficient mechanism for generating enhanced plasma biomarker protein panels. Advances in protein identification techniques for human plasma samples will likely aid our appreciation of the etiology and progression of human diseases in longitudinal studies

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Intramural Research Programs of National Institute on Aging (NIA), National Institutes of Health (NIH). We thank Dr. Luigi Ferrucci (NIA) for helpful feedback during this study and for critically reading our manuscript, and we also thank Dr. Josephine M Egan (NIA) for support during the study.

Abbreviations

DDA	data-dependent acquisition
HAPs	highly abundant proteins
MW	molecular weight
PIE	precursor ion exclusion

References

1. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics*. 2002; 1:845–867. [PubMed: 12488461]
2. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, et al. Characterization of the low molecular weight human serum proteome. *Mol. Cell Proteomics*. 2003; 2:1096–1103. [PubMed: 12917320]
3. Huttenhain R, Malmstrom J, Picotti P, Aebersold R. Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* 2009; 13:518–525. [PubMed: 19818677]

4. Polaskova V, Kapur A, Khan A, Molloy MP, et al. High-abundance protein depletion: Comparison of methods for human plasma biomarker discovery. *Electrophoresis*. 2010; 31:471–482. [PubMed: 20119956]
5. Tu C, Rudnick PA, Martinez MY, Cheek KL, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res*. 2010; 9:4982–4991. [PubMed: 20677825]
6. Whiteaker JR, Zhang H, Eng JK, Fang, et al. Head-to-head comparison of serum fractionation techniques. *J. Proteome Res*. 2007; 6:828–836. [PubMed: 17269739]
7. Musselman I, Speicher DW. Human serum and plasma proteomics. *Curr. Protoc. Protein Sci*. 2005; Chapter 24(Unit 24.1.1–24.1.18)
8. Chen HS, Rejtar T, Andreev V, Moskovets E, et al. Enhanced characterization of complex proteomic samples using LC-MALDI MS/MS: exclusion of redundant peptides from MS/MS analysis in replicate runs. *Anal. Chem*. 2005; 77:7816–7825. [PubMed: 16316193]
9. Wang N, Li L. Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem*. 2008; 80:4696–4710. [PubMed: 18479145]
10. Rudomin EL, Carr SA, Jaffe JD. Directed sample interrogation utilizing an accurate mass exclusion-based data-dependent acquisition strategy (AMEx). *J. Proteome Res*. 2009; 8:3154–3160. [PubMed: 19344186]
11. Terracciano A, Martin B, Ansari D, Tanaka T, et al. Plasma BDNF concentration, Val66Met genetic variant and depression-related personality traits. *Genes Brain Behav*. 2010; 9:512–518. [PubMed: 20345896]
12. Golden E, Emiliano A, Maudsley S, Windham BG, et al. Circulating brain-derived neurotrophic factor and indices of metabolic and cardiovascular health: data from the Baltimore Longitudinal Study of Aging. *PLoS One*. 2010; 5:e10099. [PubMed: 20404913]
13. Wu WW, Wang G, Yu MJ, Knepper MA, et al. Identification and quantification of basic and acidic proteins using solution-based two-dimensional protein fractionation and label-free or 18O-labeling mass spectrometry. *J. Proteome Res*. 2007; 6:2447–2459. [PubMed: 17506541]
14. Richter R, Schulz-Knappe P, Schrader M, Standker L, et al. Composition of the peptide fraction in human blood plasma: database of circulating human peptides. *J. Chromatogr. B Biomed. Sci. Appl*. 1999; 726:25–35. [PubMed: 10348167]
15. Yi J, Kim C, Gelfand CA. Inhibition of intrinsic proteolytic activities moderates preanalytical variability and instability of human plasma. *J. Proteome Res*. 2007; 6:1768–1781. [PubMed: 17411080]
16. Pernemalm M, Orre LM, Lengqvist J, Wikstrom P, et al. Evaluation of three principally different intact protein pre-fractionation methods for plasma biomarker discovery. *J. Proteome Res*. 2008; 7:2712–2722. [PubMed: 18549256]
17. Drake PM, Cho W, Li B, Prakobphol A, et al. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem*. 2010; 56:223–236. [PubMed: 19959616]

Clinical Relevance

Analysis of human plasma samples provides the most tractable mechanism by which longitudinal trajectories of human health can be assessed. The future design of diagnostic predictors of disease, as well as the proof of therapeutic efficacy, would strongly benefit from the development of new tools for accurate, in-depth and reliable analysis of the constituents of the plasma proteome. Unfortunately human plasma, while being one of the most important biofluids, presents a tremendous series of technical issues for mass spectrometric proteomic analysis. Chief amongst these issues are the large dynamic range of proteins present, as well as the massive over-representation of a small number of high-abundance proteins. In this study, we have implemented a precursor ion exclusion process to enhance our ability to identify peptides and proteins from highly complex plasma samples. Advances in protein identification techniques for human plasma samples will aid our appreciation of the etiology and progression of human diseases in longitudinal aging studies.

Figure 1. Box-and-whisker distributions of proteins corresponding to different MW fractions for (A) depleted plasma; and (B) HCT116 cell lysate.

Figure 2. Percentage of unique IDs gained (relative to baseline) by PIE and DDA (A) at peptide level; (B) at protein level. ND: non-depleted plasma; D: depleted plasma; No. 2–12 represent MW fractions of the depleted plasma.

Figure 3. Percentage of uniquely identified peptides gained from successive multiple PIE injections. ND: non-depleted plasma; D: depleted plasma; No.2–12 represent MW fractions of the depleted plasma.