Analytica Chimica Acta 904 (2016) 65–75

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Mining the human plasma proteome with three-dimensional strategies by high-resolution Quadrupole Orbitrap Mass Spectrometry

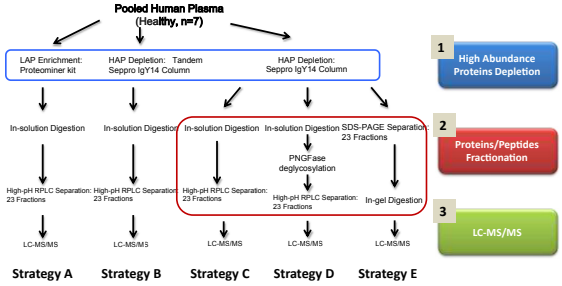
Yan Zhao [1], Cheng Chang [1], Peibin Qin, Qichen Cao, Fang Tian, Jing Jiang, Xianyu Li, Wenfeng Yu, Yunping Zhu*, Fuchu He*, Wantao Ying*, Xiaohong Qian*

State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, No. 33 Life Science Park Road, Changping District, Beijing 102206, PR China

HIGHLIGHTS

- Five three-dimensional strategies were developed and compared.
- The Strategy combined single Seppro IgY14 immunodepletion, high-pH RPLC fractionation and LC-MS/MS analysis was preferred as the optimum three-dimensional strategy.
- The optimal strategy generated 1544 plasma protein groups and 258 newly identified proteins.
- Up to 20 cytokines in the concentration range from sub-nanograms/milliliter to micrograms/milliliter were identified.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form
26 October 2015
Accepted 4 November 2015
Available online 25 November 2015

Keywords:
Plasma
Proteome profiling
IgY14 immunodepletion
Biomarker

ABSTRACT

Human plasma is a readily available clinical sample that reflects the status of the body in normal physiological and disease states. Although the wide dynamic range and immense complexity of plasma proteins are obstacles, comprehensive proteomic analysis of human plasma is necessary for biomarker discovery and further verification. Various methods such as immunodepletion, protein equalization and hyper fractionation have been applied to reduce the influence of high-abundance proteins (HAPs) and to reduce the high level of complexity. However, the depth at which the human plasma proteome has been explored in a relatively short time frame has been limited, which impedes the transfer of proteomic techniques to clinical research. Development of an optimal strategy is expected to improve the efficiency of human plasma proteome profiling.

Here, five three-dimensional strategies combining HAP depletion (the 1st dimension) and protein fractionation (the 2nd dimension), followed by LC-MS/MS analysis (the 3rd dimension) were developed and compared for human plasma proteome profiling. Pros and cons of the five strategies are discussed for two issues: HAP depletion and complexity reduction. Strategies A and B used proteome equalization and tandem Seppro IgY14 immunodepletion, respectively, as the first dimension. Proteome equalization (strategy A) was biased toward the enrichment of basic and low-molecular weight proteins and had limited ability to enrich low-abundance proteins. By tandem removal of HAPs (strategy B), the efficiency

* Corresponding authors. Tel.: +86 10 80705055; fax: +86 10 80705155.
E-mail addresses: zhuyunping@gmail.com (Y. Zhu), hefc@bmi.ac.cn (F. He), proteomics@126.com (W. Ying), qianxh1@163.com (X. Qian).
[1] These authors contributed equally to this work. The former performed all experiments, and the latter performed data processing.

http://dx.doi.org/10.1016/j.aca.2015.11.001

of HAP depletion was significantly increased, whereas more off-target proteins were subtracted simultaneously. In the comparison of complexity reduction, strategy D involved a deglycosylation step before high-pH RPLC separation. However, the increase in sequence coverage did not increase the protein number as expected. Strategy E introduced SDS-PAGE separation of proteins, and the results showed oversampling of HAPs and identification of fewer proteins. Strategy C combined single Seppro IgY14 immunodepletion, high-pH RPLC fractionation and LC-MS/MS analysis. It generated the largest dataset, containing 1544 plasma protein groups and 258 newly identified proteins in a 30-h-machine-time analysis, making it the optimum three-dimensional strategy in our study. Further analysis of the integrated data from the five strategies showed identical distribution patterns in terms of sequence features and GO functional analysis with the 1929-plasma-protein dataset, further supporting the reliability of our plasma protein identifications. The characterization of 20 cytokines in the concentration range from sub-nanograms/milliliter to micrograms/milliliter demonstrated the sensitivity of the current strategies.

1. Introduction

Human plasma transports substances between organs and tissues and carries a large number of proteins. Thus plasma can provide a large amount of physiological and pathological information. The plasma proteome is a rich resource that is thought to contain biomarkers that are useful for disease diagnosis, prognosis and treatment monitoring [1]. Relative quantification analysis of the plasma proteome has attracted the attention of many clinicians [2]. In addition, by comparing the differential proteomes of cell lines, organs and tissues, a long list of biomarker candidates was produced. If these candidates can be detected in the plasma, the translation of proteomics research into clinical applications will occur much more rapidly [3]. Thus, for either biomarker discovery [4] or further verification of biomarkers [5], broad coverage of the plasma proteome is extremely important.

Identification of biomarkers in the blood proteome is hindered by the broad concentration range of blood proteins and their complex composition [6]. Low-abundance proteins (LAPs) are those with concentrations in the nanograms/milliliter range. These proteins are known as the most promising biomarker candidates, and they include proteins that leak into the blood from tissues as well as cellular ligands and signaling molecules. LAPs tend to be obscured by high-abundance proteins (HAPs) and have a lower chance of being detected [7]. HAP depletion, prefractionation and selective capturing of subproteomes have been performed to address the complexity of the blood proteome, and various methods [7,8] such as immunodepletion [9], protein equalization [10], SDS-PAGE [11], SCX [12], IEF [13], high-pH RPLC separation [11] and multi-lectin affinity enrichment [14] have been implemented. When these methods were used alone, only small-scale datasets were produced. For example, using 20-HAPs immunodepletion or protein equalization followed by nano-LC-data-independent acquisition MS analysis [15], only approximately 272 and 200 proteins were identified from human plasma samples, respectively. Even when using serial immunodepletion in tandem to remove abundant proteins [16], the maximum number of proteins identified did not exceed 300 in human plasma samples. However, the HAPs depletion method combined with hyper-fractionation can effectively increase the depth of analysis. In our previous study, a two-dimensional peptide separation strategy coupling Offline SCX and RPLC separation was employed for proteomic analysis. A total of 529 unique proteins were identified in the flow-through and bound fractions from immunodepletion columns for the top6 and the top12 proteins in human plasma [17]. Cao et al. [11] compared the performances of SDS-PAGE, IEF and high-pH RPLC for fractionation of human plasma. Up to 592 proteins were identified by at least two peptides utilizing high-pH RPLC as the middle step in immunodepletion plasma proteome profiling. Millioni et al. [18] used a four-dimensional method (proteome equalization-IEF-SCX-RPLC-MS) to increase the number of fractions to 32 and identified 583 proteins.

Although these strategies, which have been used by various research groups, have greatly increased the amount of plasma proteome data, the number of plasma proteins identified by a single analysis is much smaller than the total number of plasma proteins. According to data updated in 2014, 10,546 proteins linked to 509 scientific articles were listed in the Plasma Proteome Database (http://www.plasmaproteomedatabase.org) [19]. The development of a strategy for obtaining a greater depth of plasma proteome profiling in a short time is still urgently needed and is a continuous pursuit of clinical and basic research.

The utilization of fast, high-resolution mass spectrometry will increase the speed of protein identification. More than 2500 proteins can be identified in standard 90-min gradients in mammalian cell lysates using Q-Exactive [20]. Thus, in this study, via High-resolution Quadrupole Orbitrap Mass Spectrometry analysis, different combinations of separation techniques were evaluated. The features and performance of five strategies coupling HAP depletion and complexity reduction methods were compared. Based on an adequate understanding of the characteristics of different separation techniques, an optimal multidimensional separation strategy will be presented that can improve the “bucket effect” of the methods and allow in-depth identification of human blood proteins with greater efficiency.

2. Materials and methods

2.1. Samples and reagents

Human plasma samples were supplied by the China Human Liver Proteome Project (CNHLPP) sample bank. Seven healthy plasma samples were pooled and prepared for the subsequent experiment. Trypsin (sequencing grade) was obtained from Promega (Madison, WI). Unless otherwise noted, all other reagents were purchased from Sigma, USA.

2.2. HAPs immunodepletion and LAPs enrichment

The Seppro IgY14 LC-2 column (Sigma, USA) was used to bind human serum HSA, IgG, fibrinogen, transferrin, IgA, IgM, haptoglobin, alpha2-macroglobulin, alpha1-acid glycoprotein, alpha1-antitrypsin, Apo A-I HDL, Apo A-II HDL, complement C3 and LDL (ApoB). In accordance with the manufacturer's recommendations, a 50-μl aliquot of a crude plasma sample was diluted with Tris-