JOURNAL OF PROTEOMICS 119 (2015) 90–99

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

A comprehensive map and functional annotation of the normal human cerebrospinal fluid proteome

Yang Zhang[a,1], Zhengguang Guo[b,1], Lili Zou[b], Yehong Yang[b], Liwei Zhang[a], Nan Ji[a], Chen Shao[c], Wei Sun[b,*], Yajie Wang[d,e,**]

[a]Department of Neurosurgery, China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, 6 Tian Tan Xi Li, Beijing 100050, China
[b]Core Facility of Instrument, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, Beijing 100005, China
[c]National Key Laboratory of Medical Molecular Biology, Department of Physiology and Pathophysiology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, Beijing 100005, China
[d]Core Laboratory for Clinical Medical Research, Beijing Tiantan Hospital, Capital Medical University, 6 Tian Tan Xi Li, Beijing 100050, China
[e]Department of Clinical Laboratory Diagnosis, Beijing Tiantan Hospital, Capital Medical University, 6 Tian Tan Xi Li, Beijing 100050, China

ARTICLE INFO

Article history:
Received 25 October 2014
Accepted 13 January 2015
Available online 7 February 2015

Keywords:
Cerebrospinal fluid
Proteome
Functional annotation

ABSTRACT

Knowledge about the normal human cerebrospinal fluid (CSF) proteome serves as a baseline reference for CSF biomarker discovery and provides insight into CSF physiology. In this study, high-pH reverse-phase liquid chromatography (hp-RPLC) was first integrated with a TripleTOF 5600 mass spectrometer to comprehensively profile the normal CSF proteome. A total of 49,836 unique peptides and 3256 non-redundant proteins were identified. To obtain high-confidence results, 2513 proteins with at least 2 unique peptides were further selected as bona fide CSF proteins. Nearly 30% of the identified CSF proteins have not been previously reported in the normal CSF proteome. More than 25% of the CSF proteins were components of CNS cell microenvironments, and network analyses indicated their roles in the pathogenesis of neurological diseases. The top canonical pathway in which the CSF proteins participated was axon guidance signaling. More than one third of the CSF proteins (788 proteins) were related to neurological diseases, and these proteins constitute potential CSF biomarker candidates. The mapping results can be freely downloaded at http://122.70.220.102:8088/csf/, which can be used to navigate the CSF proteome.

Biological significance
This study identified and functionally annotated the largest high-precision dataset of the CSF proteome, which offers a baseline reference for CSF biomarker discovery and reveals insight into CSF physiology.

* *Corresponding author.*
** *Correspondence to:* Y. Wang, Core Laboratory for Clinical Medical Research, Beijing Tiantan Hospital, Capital Medical University, 6 Tian Tan Xi Li, Beijing 100050, China.
E-mail addresses: zhangyang8025@yeah.net (Y. Zhang), gzg0625@sina.com (Z. Guo), sunwei1018@hotmail.com (W. Sun), tiantanwyj@aliyun.com (Y. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jprot.2015.01.017
1874-3919/

1. Introduction

Cerebrospinal fluid (CSF) is a water-like, colorless fluid that circulates within the brain ventricles and subarachnoid spaces and surrounds the parenchyma of both the brain and spinal cord. CSF is mainly formed at the choroid plexus in the brain ventricles and contains a large number of proteins that originate from the blood [1]. A small amount of CSF is also derived from the extracellular fluid of brain tissue, which contains proteins released by cells of the central nervous system (CNS) [1]. These proteins may be altered in disease states and are believed to be a rich source of neurological disease biomarkers.

Proteomic technology has proven to be a useful tool for the discovery of biomarkers. Several proteomic studies have attempted to identify biomarkers in CSF [2–5]. However, to facilitate the process of biomarker identification, it is also important to study the normal CSF proteome because knowledge of the normal protein composition and concentration provides a baseline for comparison. Meanwhile, knowledge of the normal CSF proteome provides insight into CSF physiology, as the proteins are directly involved in physiological processes.

One of the earliest studies of the CSF proteome was led by Jos Raymackers in 2000 [6]. This work first combined mass spectrometry and 2D-PAGE to comprehensively study protein composition in the CSF. However, the proteins identified in the study do not represent a truly normal CSF proteome because the CSF samples were pooled from patients with neurological diseases, including dementia, multiple sclerosis, and suspected CNS infection. The earliest study of the normal CSF proteome, to the best of our knowledge, was conducted by Albert Sickmann et al. in 2002 [7]. In this study, 85 proteins were identified from more than 480 spots in 2D-PAGE. Because a few high-abundance CSF proteins constitute a very large fraction of the total protein amount and the range of protein concentrations spans up to 12 orders of magnitude [8,9], it is very important to deplete these high-abundance proteins or to enrich relatively low-abundance proteins before comprehensive mapping of the CSF proteome. Immunoaffinity depletion of high-abundance proteins [10–12] and lectin affinity of glycoproteins [13] have been demonstrated to be efficient methods for increasing the detection of low-abundance proteins in normal CSF. The use of a combinatorial peptide ligand library has been demonstrated to be a powerful method for reducing the dynamic range of protein concentration in the CSF, and in 2010, Mouton-Barbosa E. et al. used this method to identify a record number of proteins (1212 proteins) [14].

Recently, a few large-scale studies of the normal CSF proteome have been conducted. Steven E. Schutzer et al. [10] combined SCX-RP-HPLC with LTQ or LTQ-Obitrap to comprehensively profile CSF proteins from healthy volunteers and identified a total of 2630 non-redundant proteins; of these, 1506 had at least 2 unique peptide identifications. Guldbrandsen et al. [15] reported the largest dataset of the normal CSF proteome to date. In that study, the human cerebrospinal fluid (CSF) proteome was mapped using three different strategies prior to Orbitrap LC–MS/MS analysis: SDS-PAGE and mixed mode reversed phase-anion exchange for mapping the global CSF proteome, and hydrazide-based glycopeptide capture for mapping glycopeptides [15]. As a result, the group identified 2875 protein groups, of which 1944 had at least two unique peptide identifications. Their mapping results can be freely downloaded from an online database: the CSF Proteome Resource (http://probe.uib.no/csf-pr).

The recent rapid development of LC–MS/MS technologies has provided a powerful discovery platform that allows for the global proteomic coverage of CSF proteins. High-pH reverse-phase liquid chromatography (hp-RPLC) is well established as a first-dimension peptide fractionation method and has been shown to perform more effectively than strong cation exchange chromatography (SCX) or off-gel electrophoresis (OGE) in qualitative and quantitative proteomic studies [16–18]. The TripleTOF 5600 system combines high resolution and mass accuracy with high rates of MS/MS acquisition, and it provides a desirable discovery platform for the in-depth profiling of complex biological mixtures [19]. In the present study, hp-RPLC was first integrated with TripleTOF 5600 to comprehensively profile the normal CSF proteome. The dataset offers a useful baseline reference for CSF biomarker discovery and provides further insight into CSF physiology.

2. Experimental procedures

2.1. Apparatuses

A TripleTOF 5600 mass spectrometer from AB Sciex (Framingham, MA, USA) and an ACQUITY UPLC system from Waters (Milford, MA, USA) were used.

2.2. Reagents

Deionized water from a MilliQ RG ultrapure water system (Millipore, Bedford, MA, USA) was used at all times. HPLC-grade acetonitrile and formic acid, ammonium bicarbonate, iodoacetamide, dithiothreitol, sequencing-grade modified trypsin, and protease-inhibitor PMSF (phenylmethanesulfonyl fluoride) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.3. CSF collection

CSF samples were collected by lumbar puncture from patients who received spinal anesthesia before non-neurological operations at Beijing Tiantan Hospital. These patients were checked by an independent medical doctor to rule out neurological diseases and recent medication use. Following collection, a subsample of each CSF sample was sent to a clinical laboratory for routine CSF diagnostics. The remaining sample was immediately centrifuged for 10 min at 2500 ×*g* to remove cellular components and subsequently aliquoted and stored at −80 °C for further analysis. A total of 14 samples from 14 individuals (7 women and 7 men, aged 24 to 55 years, with a median age of 28 years) were selected and subjected to quantitation by the Bradford method [20]. Equal protein amounts from 14 CSF samples were mixed, resulting in the pooled CSF sample for the proteomic analyses. All selected samples had normal clinical laboratory values with respect to microbiology, chemistry, and cell counts. Approval for this