Published in "PROTEOMICS doi: 10.1002/pmic.201600419, 2017"
which should be cited to refer to this work.

Separation and Identification of Mouse Brain Tissue Microproteins using Top-down Method with High Resolution Nanocapillary Liquid Chromatography Mass Spectrometry

Wenxue Li[1], Filomena Petruzziello[1], Nan Zhao[2], Huiyuan Zhao[2], Xueting Ye[3], Xiaozhe Zhang[2*] and Gregor Rainer[1]

[1]Visual Cognition Laboratory, Department of Medicine, University of Fribourg, Chemin de Musee 5, Fribourg, CH-1700, Switzerland

[2]CAS Key Laboratory of Separation of Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian, 116023.

[3]Shenyang Pharmaceutical University, Wenhua Road 103, Shenyang, China.

Abstract

Microproteins and endogenous peptides in the brain contain important substances that have critical roles in diverse biological processes, contributing to signal transduction and inter-cellular signaling. However, variability in their physical or chemical characteristics, such as molecule size, hydrophobicity and charge states, complicate the simultaneous analysis of these compounds, although this would be highly beneficial for the field of neuroscience research. Here, we present a top-down analytical method for simultaneous analysis of microproteins and endogenous peptides using high-resolution nanocapillary liquid chromatography tandem mass spectrometry. This method is detergent-free and digestion-free, which allows for extracting and preserving intact microproteins and peptides for direct

LC-MS analysis. Both higher energy collision dissociation and electron-transfer dissociation fragmentations were used in the LC-MS analysis to increase the identification rate, and bioinformatics tools ProteinGoggle and PEAKS Studio software were utilized for database search. In total, we identified 471 microproteins containing 736 proteoforms, including brain derived neurotrophic factor (BDNF) and a number of fibroblast growth factors (FGF). In addition, we identified 599 peptides containing 151 known or potential neuropeptides such as somatostatin-28 and neuropeptide Y. Our approach bridges the gap for the characterization of brain microproteins and peptides, which permits quantification of a diversity of signaling molecules for biomarker discovery or therapy diagnosis in the future.

Abbreviations

LC-MS, liquid chromatography-mass spectrometry;

LTQ, linear trap quadrupole;

HCD, higher energy collision dissociation;

ETD, Electron-transfer dissociation;

BDNF, brain derived neurotrophic factor;

FGF, fibroblast growth factors;

NGF, nerve growth factor;

FTICR, Fourier transform ion cyclotron resonance;

HIC, hydrophobic interaction chromatography;

GEFrEE, gel-eluted liquid fraction entrapment electrophoresis;

PTMs, post-translational modifications;

FDR, false discovery rate;

GRAVY, grand average of hydropathicity;

SDS, sodium dodecyl sulfate;

MIF, migration inhibitory factor;

Keywords

Microproteins, top down, peptides, high-resolution mass spectrometry, mouse brain

Microproteins in the brain encompass many important substances for brain function and development. These substances are critical for regulation and control of various physiological processes. They are a group of compounds of intermediate size (5 kD-20 kD), and include many important substances contributing to physiological signaling, such as the neuronal growth factors brain-derived neurotrophic factor (BDNF) and nerve growth factor (NGF), fibroblast growth factor (FGF), proteins such as cortexin that regulate development and synaptically-released proteins such as complexin. Microproteins often functionally interact with another type of important molecules, neuropeptides [1], to regulate the complex physiological processes of the brain [2]. Given their close functional relationship, it is highly desirable to monitor both microproteins as well as endogenous peptides for providing an in-depth understanding of the different aspects of physiological processes mediated by these two classes of molecules.

Top-down proteomics has considerable advantages for microprotein identification, allowing not only description of native modifications, but also avoiding many false-positives that tend to occur in

bottom-up peptide identification. However, current top-down proteomics mostly focuses on large protein analysis and so far has tended to pay less attention to microproteins. Because microproteins share some characteristics with endogenous peptides, including for example a generally hydrophilic nature and instability at room temperature. Thus preparation procedures developed for peptides, and especially neuropeptides, might thus be used for microprotein sample extraction. Totally, it is necessary to monitoring with sensitive analytical method to facilitate the potential biological functions discovery in the future.

Although a large number of top-down approaches have been reported, it remains a challenging task to separate proteins from real complex mixtures without pre-fractions. Many extraction methods, utilizing diverse kinds of detergent buffers in addition to acetic acid buffer as extract solution, have been previously tested. These studies confirm that detergent buffers can efficiently disrupt cell membranes and release additional proteins, but these methods are well suited only for medium to large size proteins. Since a high content of detergent is incompatible with mass spectrometry analysis, additional sample preparation procedures are generally required. For example, in the desalt and remove detergent steps with methanol/chloroform/water precipitation [3], pellets tend to prove difficult to dissolve with non-detergent solvent unless high concentration cold formic acid is used [4]. Our investigation shows that with a high power sonication step (100 kHz), acetic acid buffer alone – without the need of detergents – is in fact able for extraction of both microproteins and peptides (See Support Information Figure S1).

With the advances in the high-resolution mass spectrometry [6, 7] and search software[8], top-down proteomics for complex sample LC-MS analysis has been much improved. However, there remain some limitations for different mass spectrometry instruments that are dependent on resolve power. For example, 105 kDa proteins can be analyzed on the “12 T LTQ FT Ultra”, which possesses a resolve power of more than 200,000 [9]. Previous investigation showed that the resolve power of the high field “Orbitrap Elite” is well suited for the top-down analysis of proteins up to about 30 kDa [7]. In addition to the limitation of analyzer capability, the complexity of samples is another major limiting factor in top-down proteomics. Thus, sample fractionations are often performed prior to LC-MS

analysis, where efficient fractionation options are generally selected based on previous mass analysis, including for example GELFrEE [9-11], hydrophobic interaction chromatography combined with ion exchange, or size exclusive and affinity chromatography [12]. Here, we have tested a GELFrEE unit for microprotein fractionation with SDS detergent as lysis buffer and found that most of peptides and some microproteins were in fact not easily fractioned even using a high content gel cartridge (See Support Information Figure S2). Furthermore, most of the peptides were subsequently lost during the deleting detergent steps after GELFrEE fractionation. In comparison, acetic acid solutions plus super-filtration were found to be most suitable for the simultaneous extraction of microproteins and endogenous peptides, addressing some of the limitations of the above mentioned preparation methods. We observed that most identified microproteins extracted with acetic acid aqueous buffer as extract solution were up to 20 kDa in size (See Figure 2A). The summary of workflow has been shown at Figure 1.

In present study, 471 proteins were identified with HCD and ETD fragmentation technologies (for complete list see Support Information Table S1). As shown in Fig 2A, the molecular weight of the majority of proteins ranged between 5 kDa to 20 kDa (up to 84.90%), which corresponds to the microprotein range of the present study. Among these proteins, there were a number of known secreted proteins, which have important roles in diverse biological processes and physiological pathways [13]. Overall, secreted proteins account for approximately 21.2% (Figure 3C) of identified proteins. The Acyl-CoA-binding protein, a representative microprotein, identified using the ProteinGoggle software, is shown in Figure 3. A number of unbound brain tissue proteins, as can be released following cell membrane disruption, were also identified. We also document a number of functional microproteins, including for example brain-derived neurotrophic factor (BDNF) [14], which is active in the hippocampus, cortex, and basal forebrain areas and supports learning and higher nervous system function [15], as well as long-term memory [16]. We detected Fibroblast growth factors (FGF), signaling molecules belonging to a family that has 22 members [17], with FGF1, FGF2, FGF6 and FGF22 identified as microproteins in the present study. Among the functions of FGF is the promotion of cardiac reprogramming under defined serum-free conditions, as has recently been reported for FGF2 and FGF10 [18]. The microprotein complexin-4 was also identified here, which acts as a positive

regular of synaptic vesicle exocytosis and can either serve as a promoter or an inhibitor of vesicle fusion by influencing neurotransmission.

Among the other detected microproteins were the second messenger G protein (Guanine nucleotide binding protein) and calmodulin (calcium-modulated proteins). G proteins are commonly distributed in most cells, serving as important signal transducing molecules, whereas calmodulin is involved in numerous physiological processes including metabolism, apoptosis, short-term and long-term memory and the immune response. Our findings suggest that stability and reliability of the analytical method is of prime importance in the study of secreted as well as other unbounded proteins from brain tissue. Note that non-secreted proteins may also contribute to potential biological activity or carry information related to particular brain disorders. It worth noting that macrophage migration inhibitory factor (MIF) was also identified using our method, a microprotein that is critically involved in the anxiety, depression and memory-related behaviors [19]. In addition to a pro-inflammatory function, MIF thus also functions in a manner similar to neuropeptides making it a potentially relevant molecular target for the development of treatments linked to disorders involving anxiety and depression. The identification of a number of functional microproteins in the present study now permits future application of the top-down method for direct monitoring these substances for biological studies.

Although the bottom-up approach has proved very powerful, there are also some limitations associated with this method, particularly related to the interpretation of protein species or various protein isoforms resulting from alternative splicing. The top-down strategy on the other hand exhibits advantages in localizing multiple post-translational modifications in pure proteins or real complex samples. When considering the post-translational modifications, up to 736 proteoforms were collected in the present study. Proteoforms containing 305 phosphorylations, 78 methylations, 52 dimethylations, 10 trimethylations and 268 aceylations were detected and characterized, as shown in Figure 2B. The subcellular locations of identified microproteins were elucidated by consulting the Uniprot database and the resulting classifications for identified microproteins are illustrated in Figure 3C.

In addition to the microproteins, we identified 599 endogenous peptides in this experiment and 151 of these were known or potential neuropeptides such as for example somatostatin-28 and neuropeptides Y (See Support Information Table S2). Different fragmentation methods were compared for the peptides identification and the charged state distribution has also been summarized (See Support Information Figure S4). Different fragmentations technologies HCD and ETD have the supplementary effect for endogenous peptides identification and the ETD prefer to the high charge state ions than HCD fragmentation method, with representative MS2 spectra shown in the Support Information Figure S5. Post-translational modifications (PTMs) were also examined during peptide identification. PTMs can have important biological influence, for example increasing neuropeptides' affinity to receptors or making peptides more resistant to enzymatic degradation [2, 20]. In the present study, we characterized 114 peptides containing PTMs (see Support Information Table S2). N-terminal acetylation was the highest frequency PTM we were able to identify. In addition, we also observed C-terminal amidation, pyroglutamylation from glutamic acid (E) and glutamine (Q), disulfide bonding and dehydration PTMs. C-terminal amidation is specific to endogenous peptides and is a necessary PTM for functional activation of many neuropeptides [21, 22]. Our study identified a number of phosphorylated endogenous neuropeptides, such as for example the secreted protein secretogranin-1, which is a main precursor of a series of neuropeptides exhibiting more than twenty phosphorylation sites in the amino acids sequence. We documented approximately 36 secretogranin-1 derived peptides containing phosphorylation sites (Tyr348, Ser365 and/or Ser428) by searching with phosphorylation (STY) as variable PTM (See Support Information Figure S6). The phosphorylation site Try348 was indeed identical with the online searching UniProt database (Protein accession: P16041, SCG1_MOUSE, http://www.uniprot.org/uniprot/P16014). In addition, theSer365 and Ser428 information was also obtained from PhosphoSite Plus (http://www.phosphosite.org/proteinAction.do?id=4659). These analyses confirm the reliable identification of phosphorylated peptides in the present study despite the absence any enrichment method prior to LC-MS/MS analysis

This work presents the simultaneously separation and characterization of endogenous peptides and microproteins by the nanocapillary liquid chromatography tandem high-resolution mass spectrometry. Our sample preparation procedure demonstrates excellent compatibility for extraction of microproteins

and peptides, while preserving them in an intact state. The use of high-resolution fragmentation method in both ETD and HCD modes is shown to be beneficial for the simultaneous monitoring. The new algorithm search engine ProteinGoggle, which uses the isotopic mass-to-charge ratio and envelope fingerprinting for the spectrum interpretation, demonstrates excellent identification capability. The workflow presented here could be applied to monitoring of known endogenous peptides and microproteins, but is also suitable for identification and discovery of novel peptides and microproteins with as yet unknown biological functions. However, a limitation of this method for simultaneous analysis of peptides and proteins should be considered: due to the coexistence of proteins and peptides, which increases sample complexity, the determination of low abundance microproteins and endogenous peptides is limited. In spite of this, our results represent a promising choice for co-monitoring different kinds of low molecule weight proteins and peptides for biomarker discovery or therapy diagnosis in the future.

Acknowledgments

We thank Prof. ZhixinTian and Dr. Kaijie Xiao for supporting in the ProteinGoggle software application. This work was supported by project NSF 21475128 MOST 2015DFG42460 to XZ.

The authors declare no competing financial interest.

Supporting Information

The method related information at the Supplementary Information and the identified microptoteins and endogenous peptides in the Table S1 and Table S2, respectively.

Corresponding Author

E-mail: zhangxz@dicp.ac.cn. Phone: +86-411-84379667.

References

[1] Hokfelt, T., Bartfai, T., Bloom, F., Neuropeptides: opportunities for drug discovery. *Lancet Neurol* 2003, *2*, 463-472.

[2] Hokfelt, T., Broberger, C., Xu, Z. Q., Sergeyev, V., *et al.*, Neuropeptides--an overview. *Neuropharmacology* 2000, *39*, 1337-1356.

[3] Wessel, D., Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* 1984, *138*, 141-143.

[4] Doucette, A. A., Vieira, D. B., Orton, D. J., Wall, M. J., Resolubilization of precipitated intact membrane proteins with cold formic acid for analysis by mass spectrometry. *J Proteome Res* 2014, *13*, 6001-6012.

[5] Regnier, F. E., Julka, S., Primary amine coding as a path to comparative proteomics. *Proteomics* 2006, *6*, 3968-3979.

[6] Breuker, K., Jin, M., Han, X., Jiang, H., McLafferty, F. W., Top-down identification and characterization of biomolecules by mass spectrometry. *Journal of the American Society for Mass Spectrometry* 2008, *19*, 1045-1053.

[7] Ahlf, D. R., Compton, P. D., Tran, J. C., Early, B. P., *et al.*, Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J Proteome Res* 2012, *11*, 4308-4314.

[8] Li, L., Tian, Z., Interpreting raw biological mass spectra using isotopic mass-to-charge ratio and envelope fingerprinting. *Rapid communications in mass spectrometry : RCM* 2013, *27*, 1267-1277.

[9] Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., *et al.*, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, *480*, 254-258.

[10] Tran, J. C., Doucette, A. A., Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal Chem* 2008, *80*, 1568-1573.

[11] Vellaichamy, A., Tran, J. C., Catherman, A. D., Lee, J. E., *et al.*, Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal Chem* 2010, *82*, 1234-1244.

[12] Xiu, L., Valeja, S. G., Alpert, A. J., Jin, S., Ge, Y., Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal Chem* 2014, *86*, 7899-7906.

[13] Hathout, Y., Approaches to the study of the cell secretome. *Expert Rev Proteomics* 2007, *4*, 239-248.

[14] Acheson, A., Conover, J. C., Fandl, J. P., DeChiara, T. M.*, et al.*, A BDNF autocrine loop in adult sensory neurons prevents cell death. *Nature* 1995, *374*, 450-453.

[15] Yamada, K., Nabeshima, T., Brain-derived neurotrophic factor/TrkB signaling in memory processes. *Journal of pharmacological sciences* 2003, *91*, 267-270.

[16] Bekinschtein, P., Cammarota, M., Katche, C., Slipczuk, L.*, et al.*, BDNF is essential to promote persistence of long-term memory storage. *Proceedings of the National Academy of Sciences of the United States of America* 2008, *105*, 2711-2716.

[17] Blaber, M., DiSalvo, J., Thomas, K. A., X-ray crystal structure of human acidic fibroblast growth factor. *Biochemistry* 1996, *35*, 2086-2094.

[18] Yamakawa, H., Muraoka, N., Miyamoto, K., Sadahiro, T.*, et al.*, Fibroblast Growth Factors and Vascular Endothelial Growth Factor Promote Cardiac Reprogramming under Defined Conditions. *Stem cell reports* 2015.

[19] Conboy, L., Varea, E., Castro, J. E., Sakouhi-Ouertatani, H.*, et al.*, Macrophage migration inhibitory factor is critically involved in basal and fluoxetine-stimulated adult hippocampal cell proliferation and in anxiety, depression, and memory-related behaviors. *Molecular psychiatry* 2011, *16*, 533-547.

[20] Petruzziello, F., Fouillen, L., Wadensten, H., Kretz, R.*, et al.*, Extensive characterization of Tupaia belangeri neuropeptidome using an integrated mass spectrometric approach. *J Proteome Res* 2012, *11*, 886-896.

[21] Cuttitta, F., Peptide amidation: signature of bioactivity. *The Anatomical record* 1993, *236*, 87-93, 172-173; discussion 193-175.

[22] Eipper, B. A., Stoffers, D. A., Mains, R. E., The biosynthesis of neuropeptides: peptide alpha-amidation. *Annual review of neuroscience* 1992, *15*, 57-85.

Legends

Figure 1 The workflow for the microproteins and endogenous peptides analysis.

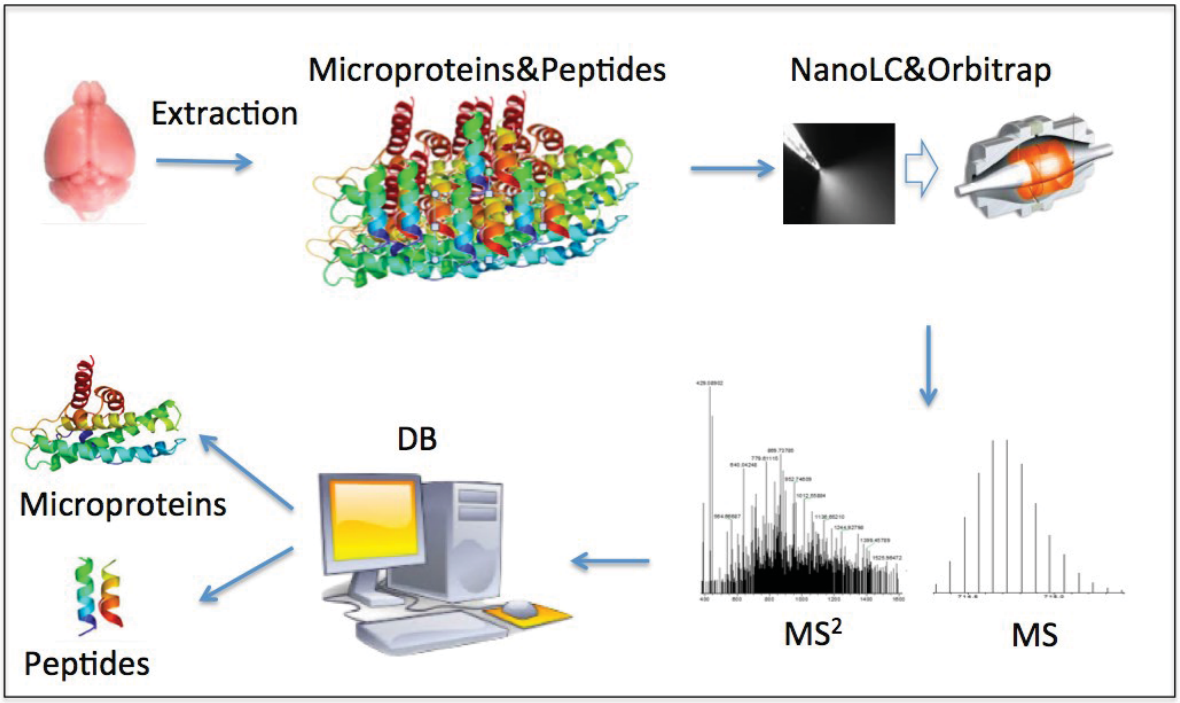

Figure 2 The scheme and the representative chromatography. A) The scheme; B) The base peak model chromatogram for microproteins and peptides by high-resolution mass spectrometry.

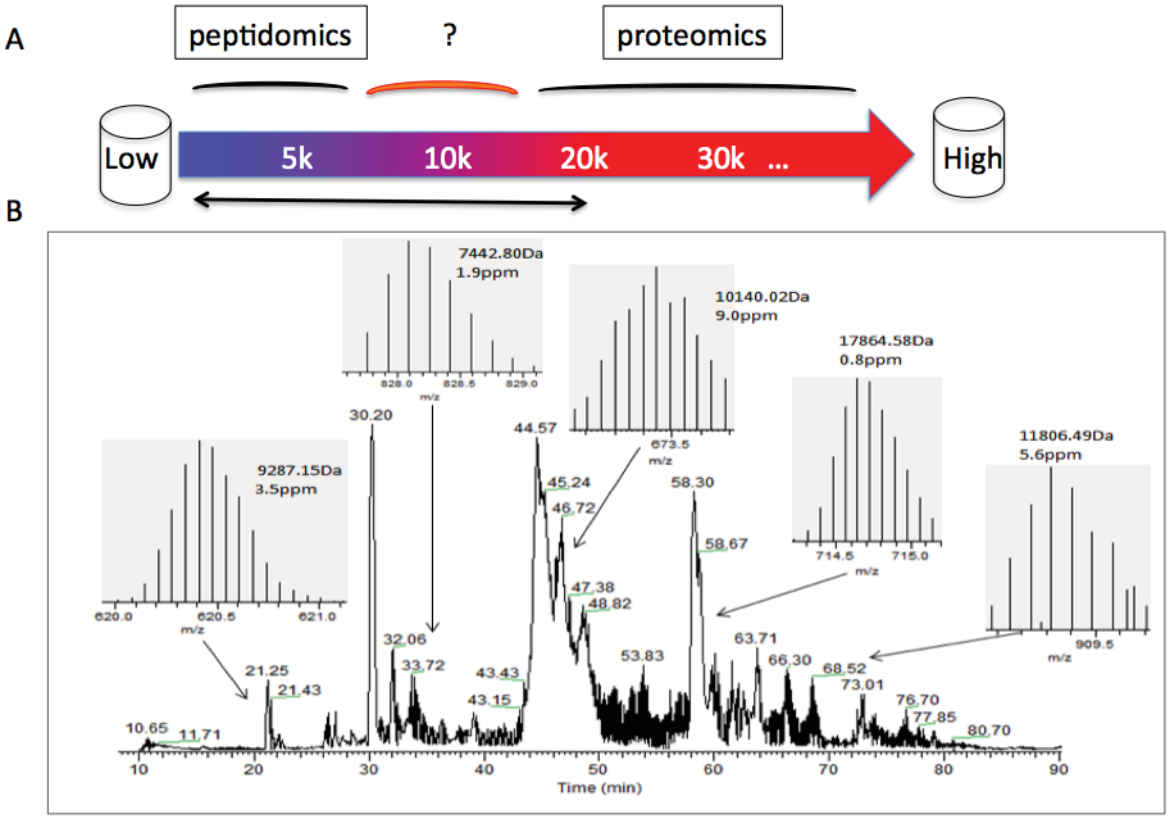

Figure 3 Identified microproteins from the LC-MS/MS analysis.A) The molecular weight distribution of proteins in the ETD and HCD;B) The PTMs distribution of identified proteoforms; C) The subcellular location of identified proteins; D) The GRAVY index distribution of identified proteins. The GRAVY index was calculated with online GRAVY calculator: http://www.gravy-calculator.de.