Published in final edited form as:

Anal Chem. 2017 November 21; 89(22): 12030–12038. doi:10.1021/acs.analchem.7b02426.

Continuous Elution Proteoform Identification of Myelin Basic Protein by Superficially Porous Reversed-Phase Liquid Chromatography and Fourier Transform Mass Spectrometry

Daniel A. Plymire[†], Casey E. Wing[‡], Dana E. Robinson[‡], and Steven M. Patrie^{†,‡,*}

[†]Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States

[‡]Department of Pathology, UT Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, United States

Abstract

Myelin basic protein (MBP) plays an important structural and functional role in the neuronal myelin sheath. Translated MBP exhibits extreme microheterogeneity with numerous alternative splice variants (ASVs) and post-translational modifications (PTMs) reportedly tied to central nervous system maturation, myelin stability, and the pathobiology of various de- and dys-myelinating disorders. Conventional bioanalytical tools cannot efficiently examine ASV and PTM events simultaneously, which limits understanding of the role of MBP microheterogeneity in human physiology and disease. To address this need, we report on a top-down proteomics pipeline that combines superficially porous reversed-phase liquid chromatography (SPLC), Fourier transform mass spectrometry (FTMS), data-independent acquisition (DIA) with nozzle-skimmer dissociation (NSD), and aligned data processing resources to rapidly characterize abundant MBP proteoforms within murine tissue. The three-tier proteoform identification and characterization workflow resolved four known MBP ASVs and hundreds of differentially modified states from a single 90 min SPLC-FTMS run on ~0.5 μ g of material. This included 323 proteoforms for the 14.1 kDa ASV alone. We also identified two novel ASVs from an alternative transcriptional start site (ATSS) of the MBP gene as well as a never before characterized S-acylation event linking palmitic acid, oleic acid, and stearic acid at C78 of the 17.125 kDa ASV.

Graphical Abstract

*Corresponding Author: steven.patrie@northwestern.edu; Phone: 847-491-3731.

ORCID

Casey E. Wing: 0000-0002-6437-0129

Steven M. Patrie: 0000-0002-0308-5498

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b02426.

Additional information as noted in the text includes supporting methods and workflow diagrams (PDF) Non-MBP proteins ID; list of MBP 14.1 ID (XLSX)

The myelin sheath, assembled by oligodendrocytes in the central nervous system (CNS) and Schwann cells in the peripheral nervous system (PNS), serves an essential role in the saltatory conduction along myelinated axons.^{1–3} Abnormal myelin development or destruction is observed in dysmyelinating white matter disorders and demyelinating diseases such as multiple sclerosis (MS).^{1,4} Myelin basic protein (MBP), an abundant myelin constituent, is a positively charged protein that is believed to stabilize the sheath's negatively charged phospholipid bilayers, in addition to other roles in neuronal signaling, cytoskeleton stabilization, and regulation of gene transcriptional machinery.^{2,5,6} MBP's functional diversity is believed to be related to its significant structural diversity associated with long regions of disorder, numerous alternative splice variants (ASVs) (10–30 kDa), and post-translational modifications (PTMs).^{1,7,8} Occurrence of “classical” MBP ASVs (~13–22 kDa) varies by species, with the 14.1 kDa ASV predominant in adult rodents and the 18.4 kDa ASV most abundant in humans.⁷ The functions of the different ASVs are not well understood, but their expression is dynamic with Golli-MBP ASVs expressed during embryonic development, followed by 17.125, 17.140, and 20 kDa ASVs in maturing animals, while the 14.1 and 18.4 kDa ASVs are dominant during adulthood.⁹ With regards to MBP PTMs, diverse modification classes have been reported (e.g., methylation (METH), N-terminal acetylation (N-ACET), deimination (citrullination), deamidation, phosphorylation (PHOS), methionine sulfoxide (MSO), and sulfone (MSONE)), many of which are believed to regulate its functional roles, such as influencing electrostatic interactions within the phospholipid bilayer and regulating signaling cascades through protein interactions.^{7,8,10–12} PTMs have been characterized in bovine, rabbit, spiny dogfish, chicken, rattlesnake, and human MBP through peptide-based proteomics, identifying over 40 putative sites on the 18.4 kDa ASV.^{8,13–17} This work, and others, has begun to reveal intriguing insights on how MBP PTM status correlates with autoimmune activities in MS and corresponding animal models (e.g., experimental autoimmune encephalomyelitis (EAE)).¹⁸

Despite decades of research, there is still a limited understanding of how this intrinsically disordered protein utilizes chemical-combinatorial space that derives from both ASVs and PTMs to regulate its structure and function in CNS and PNS development, homeostasis, and dysfunction.¹¹ This is largely because conventional bioanalytical approaches (e.g., Western blots, gel electrophoresis, and peptide-based liquid chromatography–mass spectrometry (LCMS)) poorly characterize extreme proteoform microheterogeneity for a single protein. Considering that animal models are commonly used in physiology, immunology, and

neuroscience studies into the importance of myelin in health and disease and that significant species differences in MBP ASVs and PTMs exist, it is critical to have technologies that can rapidly separate, identify, and quantify the MBP proteoform landscape in a standardized and sensitive screening environment. Such resources may help decipher unique transcriptional or post-translational processes during disease pathobiology as well as provide insights on how species differences may adversely impact translation of new treatments from rodent models to human patients.¹⁹ To address this need, this study reports on new top-down mass spectrometry (TDMS)²⁰ resources that prove effective for discrete proteoform characterization on MBP isolated from CNS tissues (i.e., brain and spine). TDMS typically employs LC, intact mass analysis, and gas-phase fragmentation (e.g., tandem mass spectrometry, MS/MS) to characterize protein's sequences, localize the positions of their PTMs, and quantify related proteoform ratios.^{21–23} While applied in comprehensive proteomics environments that identify thousands of proteins in samples,²⁴ TDMS has proven especially valuable for in-depth proteoform investigations (e.g., histones, glycoproteins, cardiac proteins, and virulence factors^{22,25–30}) that are aided by the high resolving power and mass accuracy afforded by Fourier transform mass spectrometry (FTMS),^{31,32} as well as offline and online data-dependent (DD) fragmentation, diverse 1D and 2D LC techniques, and specialized bioinformatics tools.^{20,35,36} Herein, we highlight a new workflow that combines superficially porous reversed-phase liquid chromatography (SPLC) and fragmentation by data-independent acquisition (DIA) on a FTMS with a custom bioinformatics workflow^{31,35} for the rapid interrogation of the MBP proteoform landscape. Of note, along with separating most MBP ASVs from one another, SPLC also separated numerous differentially modified proteoforms of each ASV, thus improving the discrimination of proteoforms with the same chemical composition but with the PTM localized to different amino acids (i.e., isobars). To manage the identification of the extreme MBP microheterogeneity we postulated, compared with DD methods where species are fragmented sequentially, parallel fragmentation with DIA would effectively identify coeluting proteoforms at varied abundance over the course of the LC run. The three-tier absolute mass search strategy implemented with continuous elution probability-based scoring permitted automated discovery of numerous classical MBP variants and hundreds of proteoforms, which permitted more time for manual discovery of novel ASVs and PTMs. This work contributes to better understanding of MBP proteoform complexity as well as the emerging field of proteoform measurement science, which seeks to tackle the breadth of proteome space associated with a single protein.

MATERIALS AND METHODS

Isolation and Purification of MBP

Solvents and acids were Optima grade (Thermo Fisher Scientific, Waltham, MA). Animal protocols were approved by the UTSW Institutional Animal Care and Research Advisory Committee. C57BL/6 mice were anesthetized and perfused with phosphate-buffered saline prior to tissue extraction. Tissue was flash frozen and stored at -80°C . MBP extraction was accomplished via the Folch/acid protocol.³³

Mass Spectrometry and Data Analysis

A nanospray SPLC column with a 15 μm electrospray (ESI) tip, 75 μm ID, and 360 μm OD (New Objective, Woburn, MA) was packed with C18 Poroshell-300 resin 5 μm in diameter with 300 Å pores (Agilent, Santa Clara, CA) to a length of 15 cm. Approximately, 0.5 μg of MBP was separated with a 90 min linear elution gradient on an 1100 nano-LC (Agilent, Santa Clara, CA). The mobile phase was composed of 0.025% TFA and 0.3% formic acid in 5:95 acetonitrile (ACN)/water (solvent A) and 0.025% TFA and 0.3% formic acid in 80:20 ACN/isopropanol (solvent B). The flow rate was 0.5 $\mu\text{L}/\text{min}$, and the gradient was from 0 to 30% solvent B over 90 min. The column was interfaced to an LTQ Orbitrap XL/ETD (Thermo Fisher, Pittsburgh, PA) with a custom ESI source. The mass spectrometer was tuned for the $[\text{M} + 10\text{H}]^{10+}$ of ubiquitin by direct infusion. For intact mass studies, samples were analyzed at a resolving power (RP) of 60 000 at 400 m/z and nozzle skimmer dissociation (NSD) potential of 25 V. For DIA studies, samples were analyzed at a RP of 60 000, NSD potential of 60 V, and a capillary temperature of 375 °C. For both intact and DIA studies, the maximum accumulation time was 500 ms, and two μscans were collected.

Database Generation and Data Analysis

A three-tiered absolute mass search strategy was employed for protein and proteoform identifications (Figure S-2). The searches utilized intact masses and fragment data collected with DIA where all proteins/proteoforms eluting at a given elution period are fragmented in parallel using NSD. The first tier (Tier 1) was against the *Mus musculus* proteome and designed to automatically differentiate known MBP ASVs from non-MBP proteins across the elution period. Automated searches in the second tier (Tier 2) were made against combinatorially annotated³⁴ MBP-ASV/PTM databases to differentiate the presence of unique MBP proteoforms. The third tier (Tier 3) was performed manually and used the *m* mode search in ProSightPC 3.0 (Thermo Fisher Scientific, Waltham, MA) to characterize unknown sequence variants and modifications.³⁵

Data Deconvolution—The AutoXtract (Isotopically Resolved) algorithm in Protein Deconvolution 4.0 (PD4, Thermo Fisher, Pittsburgh, PA) was used for monoisotopic mass determination over time within .raw files. Typical deconvolution settings used were: signal-to-noise (S/N), 1.0; minimum number of detected charges for intact protein, 2; minimum number of detected charges for fragments, 1; isotopologue fit factor, 80%; isotopologue remainder threshold, 80%; monoisotopic mass merge tolerance, 15 ppm; and target average time window, 1.0 or 0.5 min. PD4 outputs were converted into observed mass, retention time, intensity, and estimated grand average hydrophobicity index (GRAVY) values, the latter determined by an internal calibration curve created from known proteins within the sample. Monoisotopic masses (denoted by the *-0* isotopologue label in figures) were consistently reported for intact proteoforms throughout the manuscript.

Database (DB) Generation—Tier 1 databases were compiled from the UniProtKB *Mus musculus* reference proteome (Proteome ID: UP000000589) downloaded 12/16/2016. DBs were generated in ProSightPC 3.0 with no more than seven sequence events per sequence with each PTM-considered variable. Gene and proteoform information were captured using an in-house MySQL script, and the theoretical isoelectric point and GRAVY data were

calculated for each proteoform with ProPAS.1.1.pl.³⁶ Each proteoform within the DB was assigned a proteoform designation that reflects the protein identifier name followed by the type of PTM and the location in parentheses. For example, P04370-8_Ac(1)_Met(104) is the identifier for the 14.1 kDa MBP ASV with N-ACET and METH at positions 1 and 104, respectively. Tier 2 DBs for MBP were similarly generated from custom flat files (Figure S-1) specific to each ASV with each PTM considered as variable with the exception of fixed N-ACET (Supporting Information (SI)). PTMs and amino acid abbreviations in text follow that described in the UniProt Knowledgebase (UniProtKB).³⁷ The PTMs selected were based on previously reported PHOS, METH, N-ACET, and MSO/MSONE sites resulting from sequence alignment with multiple organisms.¹⁷ Citrullination and deamidation were not examined in this study. For all MBP proteoforms discussed, it was assumed each exists with fixed N-ACET.

Continuous Elution Absolute Mass Searches—Protein/proteoform identification was accomplished with a custom search engine (Figures S-3 and S-4) run on a Windows 2012 Server. The software supported generation of Poisson-based P -scores ($P'(i)$)³⁸ of proteins/proteoforms at sequential time intervals across an entire SPLC-NSD-FTMS data set. The software also automated data reporting and generation of various protein, fragmentation, and proteoform feature maps. For absolute mass searches, target masses were first populated from intact masses observed in both the SPLC-NSD-FTMS data set and the separate SPLC-FTMS data set performed without NSD. DB searches started with an absolute mass test utilizing a ± 1.2 Da mass tolerance. Each candidate sequence was then tested against the fragment data at all time points simultaneously across the SPLC-NSD-FTMS run with a P -score reported at every time point i ($P(i)$). To avoid spurious matches across the entire time range, a P_{decoy} value was determined at each time point i by way of a test against the same SPLC-NSD-FTMS data performed automatically on an inverted and scrambled sequence that retains the candidate's PTMs at the same amino acid location to that of the target. Assuming that all decoy matches over time are random, we created a decoy baseline that ensured that on average $<1\%$ of the hits across all time points in the decoy series yielded significant hits (<0.01). To accomplish this, the reported $P_{\text{decoy}}(i)$ values were corrected for (divided by) the number of related proteoforms in the database for the target protein sequence (designated $P'_{\text{decoy}}(I)$). Then a 99th percent confidence value at each i (denoted $P'_{\text{decoy}}(I, \alpha=0.01)$) was determined on a 5-pt moving average P'_{decoy} over time. Finally, to test significance of each $P(i)$ hit for the candidate sequence across the time series an adjusted P -score ($P'(i)$) was determined by $P(i) - P_{\text{decoy}}(I, \alpha=0.01)$ and only $P'(i) < 0.01$ were considered true hits (Figure S-4). The method employed avoided the significant DB scoring penalty used in the determination of an expectation value (e -values) that corrects P -scores for multiple testing. e -value corrections are adversely influenced by large custom databases that contain numerous highly related proteoform sequences for unrelated proteins.

Protein/Proteoform Prospective Validation—Along with ranking proteins/proteoform identifications by $P'(i)$, sequence maps were plotted for each identification, with each hit automatically ranked from 0 to 5 by criteria that gauged the quality of the fragmentation data for supporting the presence of one or more PTMs in the identified species (Figure S-5). Ranks of 2–4 were each considered valid, having at least 2 fragments confirming any

individual PTM. A rank of 3 and 4 also met the criteria that all internal modifications were confirmed in combination by at least 1 or 2 fragments, respectively. A rank of 1 was considered a plausible identification because at least one of the PTMs in the identified proteoform was supported by only 1 fragment ion. A rank of 0 implies at least one of the PTMs in the identified proteoform was not substantiated by a fragment ion. A rank of 5 indicates that the target protein did not harbor a PTM. Tabulated outputs for all protein/proteoform identifications were grouped by theoretical monoisotopic masses that present proteoforms with a similar chemical composition together. Heat maps were used to visualize the confidence of the proteoform identification over time to make qualitative comparisons of the elution patterns of different proteoforms (Figure S-4). Intensity profiles in the heat maps reflect the relative $-\log_{10}(P')$ for hits at time points where the intact mass was detected within a ± 1.2 Da tolerance. The intact mass stipulation served to discriminate against ASVs that have significant overlap of N-terminal and C-terminal fragments yet elute at different time points.

RESULTS

Separation and Identification of MBP ASVs by Continuous Elution SPLC-NSD-FTMS

Initial assessments sought to confirm SPLC was effective for separating major MBP ASVs while simultaneously enabling data-independent NSD and FTMS analysis of their abundant modification states. The described workflow centered on application of superficially porous (SP) resins³⁹ with a capillary LC interfaced online with FTMS because early tests provided evidence that the resin separated most abundant MBP ASVs from one another as well as helped to resolve different modified forms for each ASV. For example, 0.5 μ g of protein from brain tissue isolate was separated with a 90 min linear gradient both without and with DIA by NSD. A representative total ion current (TIC) chromatogram, fragmentation, and protein feature maps (Figure 1) highlighted that the sample contains numerous protein/proteoforms with masses from 7 to 22 kDa. Absolute mass searches on abundant monoisotopic masses and corresponding fragments observed across the chromatogram identified several classical MBP ASVs (14.1, 17.125, 17.140, and 18.4 kDa) as well as other mouse proteins (e.g., ubiquitin and numerous histone variants of H1, H2A, H2B, and H4) (Figure 1C and Supplemental Table 1). Compared to non-MBP proteins, each of the MBP ASVs eluted relatively early in the SPLC run in an order consistent with GRAVY predictions made from their amino acid sequences ($14.1 < 17.125 < 17.140 < 18.4$ kDa). For adult mice, the ratios of all MBP ASVs (Figure 1C, inset) determined for each “classical” variant isolated from brain tissue was largely consistent with gel electrophoresis (Figure 1A, inset), which showed rodent MBP predominately exists as the 14.1 and 18.4 kDa ASVs.

Identification of Classical MBP Variants

Mass spectra corresponding to the peak elution period (centroid) for the four classical ASVs that were detected suggested three of four variants consisted of numerous proteoforms in varied relative abundance (Figure 2, left and middle). Amino acid sequence maps resulting from informatics searches on NSD data corresponding to the same elution period confirmed the presence of the N-ACET proteoform for each ASV observed (Figure 2, right). In each case, observed monoisotopic intact masses matched to theoretical monoisotopic intact

masses with <10 ppm mass error and fragmentation sequence coverage of ~20%. The proteoforms of the 14.1, 18.4, and 17.125 kDa ASVs each exhibited interproteoform mass differences consistent with METH, (14 Da), DIMETH, (28 Da), and PHOS (80 Da) in varied combinations. The 17.140 kDa ASV, which lacks the –KGRGL– METH site present in the other MBP ASVs, presented predominately in an N-ACET form (Figure 2D); however, low abundance and elution between the abundant 14.1 kDa and 18.4 kDa ASVs often obscured detection of proteoform groups associated with both the 17.125 and 17.140 kDa ASVs.

Evaluation of 14.1 KDa ASV PTM Microheterogeneity

We next examined the elution profiles of the 14.1 kDa ASV to determine the extent that PTM heterogeneity contributed to peak broadening observed in the TIC (Figure 1A). A proteoform feature map (Figure 3A) and mass spectra of the $[M+15H]^{+15}$ charge state at six distinct time intervals (Figure 3B–G) highlighted intensity fluctuations of 13 groups of apparent nominal mass isobars over time (labeled a–m) (Figure 3A, bar inset). Inspection of interproteoform mass differences (Figure 3B–G, bar insets) relative to the N-ACET proteoform (a) showed a mixture of MSO and MSONE early in the elution period (e.g., Figure 3B–D) followed by elution of the N-ACET proteoform (Figure 3E), then the METH (14 Da, Figure 3F), and then DIMETH proteoforms (28 Da, Figure 3G). While PHOS (80 Da) without and with METH (94 Da) and DIMETH (108 Da) was generally observed throughout the elution period, DIPHOS proteoforms without and with METH and DIMETH (e.g., 160, 174, and 188 Da) as well as TRIPHOS proteoforms (not shown) were typically detectable at the end of the elution period. Evaluation of specific NSD fragments over time confirmed inferences made at the intact level (Figure 3B–G, insets). For example, the presence of a M19 and M124 MSO was observed early in the elution period on the b81 fragment (left inset) and the y43 fragment (right inset), respectively. The y43 fragment also confirmed the N-ACET proteoform eluted prior to the METH and then DIMETH proteoforms (Figure 3E,F insets), highlighting that increased methylation content increased hydrophobicity of the 14 kDa ASV.

To validate the elution characteristics of specific 14.1 kDa proteoforms for each apparent nominal mass in the groups observed, we performed a “Tier 2” search and tabulated and visualized proteoform identifications from the continuous elution NSD data over time (Figure 4 and Supplemental Table 2). In one run, a total of 323 unique proteoforms were identified for the 14.1 kDa ASV from a single SPLC-NSD-FTMS run. These proteoforms were clustered into 41 different groups by their unique theoretical intact masses and common chemical composition resulting from differing positions of PTMs along the amino acid backbone. Of these, 71 were ranked from 2 to 4 with 2 or more fragments confirming the presence of each internal PTM. Inspection of averaged adjusted *P*-scores (Figure 4A) for all identified proteoforms associated with each apparent nominal mass series in groups a–m confirmed that the most confident matches early in the run were for MSO and MSONE forms of the 14.1 kDa ASV. These were subsequently followed by that of the N-ACET proteoform with no other modifications, then the METH, and DIMETH proteoforms. Contributions of specific proteoforms to each nominal mass (Figure 4B,C) also helps reveal specific sites of modifications as well as their differential elution characteristics. For

example, the nominal mass series (Figure 4B) associated with the MSO (14 129 Da), MSO +METH (14 143 Da), and MSONE (14 145 Da) show that the MSONE forms elute first before the MSO proteoforms. The data also shows MSO-M124 proteoforms, regardless of METH status, tend to elute after the MSO-M19 proteoforms, an observation corroborated by the b81 and y43 fragment ion data (Figure 3C,D) as well as corroborated in separate replicate runs (not shown). The data also reveal that METH (14 127 Da) and DIMETH (14 141 Da) occur predominately at R104 and to a much lesser extent at R41 or R47 for the 14 kDa ASV. Corroboration of METH-R104 status is conclusively supported by >20 b/y ions while only 2–4 b/y ions typically confirm existence of METH-R41 or METH-R47. These observations contrast the reported METH sites listed in UniProtKB PTM/Processing for mouse MBP (P04370) but are consistent with METH sites described for rat MBP (P02688). Finally, while our database did not consist of an exhaustive list of all PHOS sites, the NSD results indicate that PHOS was readily detected toward the C-terminal end of the protein as opposed to the N-terminus. However, significant phosphate neutral loss associated with the NSD method warrants prospective examination of MBP PHOS sites via alternative fragmentation methods such as electron capture/transfer dissociation (ECD or ETD).⁴⁰

Novel MBP ASVS

The initial absolute mass searches against the mouse proteome revealed several components that did not immediately translate to confident identifications. For example, many putative MBP degradation products eluted early in the gradient and were excluded from subsequent analysis. Other proteoform groups with masses of approximately 12.1 and 16.4 kDa coeluted with the 14.1 and 18.4 kDa ASVs, respectively (Figure 1C). In both cases, we treated these as novel MBP ASVs, because their masses were not consistent with predictable degradation products of the classical variants, and mass spectra for each candidate revealed proteoform microheterogeneity (i.e., observation of N-ACET, METH, and PHOS combinations) that was consistent to that of the 14.1 and 18.4 kDa ASVs (Figure 5A,B vs Figure 2A,B, middle). In both cases, the lowest mass proteoform for the 12.1 and 16.4 kDa proteoform groups had a –1961.04-0 Da mass difference relative to the 14.1 and 18.4 kDa ASVs, respectively. To localize the sequence discrepancies, we searched NSD data using the *m* mode search in ProSightPC 3.0 against the N-ACET form of each classical ASV. For the candidate variants, the sequence discrepancy was localized to the N-terminus of the 14.1 and 18.4 kDa ASVs, respectively (Figure 5A,B), suggesting both the 12.1 and 16.4 kDa variants derive from the alternative transcriptional start site (ATSS) that begins at M19 along with N-ACET. Sequences with aspartic acid at position 2 often do not have their initiator M cleaved and exhibit methionine N-ACET.^{41–43}

Defining Novel MBP Post-Translational Modifications

Another suspected MBP proteoform group at ~17.4 kDa was observed to elute after the 18.4 kDa ASV (Figure 1C). As above, the mass spectrum showed proteoform microheterogeneity and ratio data consistent to that of the 14.1 and 18.4 kDa ASVs (Figure 6A). In this case, no degradation products, ATSS event, or combinations of previously reported classical MBP exons yielded a product that matched the monoisotopic mass of the base proteoform observed (17 364.01-0 Da). Subsequent tests of the NSD data showed agreement of this species to that of the N-ACET 17.125 kDa ASV with the mass discrepancy localized to the

C78 (Figure 6A). A 238.23 Da mass difference localized to this region suggests this cysteine-containing protein (the only classical MBP ASV containing a C) is subjected to the cysteine side chain S-acylation with palmitic acid (16:0). Addition of the hydrophobic fatty acid group helps explain the significant deviation in its observed retention time relative to the unmodified form. Accurate mass information was also used to assign similar lower abundant S-acylation products of 17.125 kDa ASV that are chromatographically resolved throughout the run (e.g., oleic acid (18:2), stearic acid (18:0), etc.) (Figure 6B). While not previously reported for MBP, observation of lipidated products is consistent with proteolipid protein (PLP), another abundant myelin membrane protein that is subject to extensive S-acylation by the same fatty acids.⁴⁴

DISCUSSION

To support prospective studies on how MBP utilizes its unique chemical-combinatorial space that derives from both ASVs and PTMs to regulate its structure and function, we developed a DIA TDMS workflow that enabled interrogation of MBP ASVs and abundant PTM states simultaneously. The described workflow showed significant potential at proteoform discovery with an information content superior to antibody- or peptide-based assays yet on a similar time scale and similar sensitivity. SPLC not only resolved many MBP ASVs but also helped resolve numerous discrete proteoforms. The most abundant proteoforms were identified with a custom continuous elution informatics procedure that enabled DIA interrogation of NSD data sets. Within a single SPLC-NSD-FTMS analysis, we successfully identified four classical MBP ASVs, and with a gradient optimized for the 14.1 kDa ASV, we characterized 323 unique proteoforms that could be assigned to 41 different nominal mass groups with unique chemical compositions associated with varied PTM location along the amino acid backbone. The most common modifications detected include N-ACET, METH-R104, DIMETH-R104, C-terminal PHOS, as well as differential oxidation at either M19 or M124. The automation of Tier 1 and Tier 2 informatics steps also facilitated manual discovery of two novel ATSS variants as well as fatty acid cysteine acylation that was unique to the 17.125 kDa ASV. While the bioinformatics resources enabled identification of the most abundant proteoforms, data-dependent CID, ETD, ECD, HCD, or UVPD could yield more identifications of low abundant proteoforms and reduce ambiguity caused by a fragment that can be assigned to multiple coeluting proteoforms.²⁰ In future work, DD may particularly benefit identification of the MBP degradation products, which may result from disease processes (e.g., demyelination). Also, targeted MS/MS with ECD or ETD may prevent PTM neutral loss (e.g., phosphate or lipids) observed with NSD.

Superficially porous resins are available in several configurations,^{45–48} offering benefits of mass transfer efficiencies similar to nonporous silica while maintaining high loading capacities at back pressures consistent with standard HPLC. The optimization of the LC gradient for the non-14.1 kDa ASVs is expected to improve the number of identifications with the current resin, but new SPLC resins as well as alternative resins (e.g., monoliths,^{49,50} submicron silica particles^{51,52}) or LC strategies (e.g., IEC, HILIC, and UPLC)⁵³ are expected to provide enhanced separations and extend the platform to much larger, heavily modified proteins.

Along with continued top-down resource development, future work will use the tools to examine the MBP proteoform landscape in myelin development and in dysmyelinating white matter disorders and demyelinating diseases. Myelin development has been reported to begin in the hindbrain and proceed rostrally with those related to nursing developing earliest, followed by motor and sensory development and then learning areas.⁵⁴ It is plausible that both ASV expression and PTM expression of MBP are dynamic during the stages of development. Assessment of changes in the degeneration and recovery in the rodent EAE model may also aid in discovering changes in MBP proteoform expression that may correlate with the progression and remission of MS in humans. To ensure success in clinical proteomics studies, the quantitative fidelity of the heat map intensity needs to be examined. The intensity metric based on proteoform ID confidence (*P*-score) is dependent upon the number of fragments detected and may correlate to proteoform abundance; however, factors such as differential ionization efficiency (e.g., different ASVs), PTM dependent fragmentation tendencies, and location of a PTM along the protein backbone may interfere with accurate proteoform quantitation.

Validation of assigned proteoforms, particularly those of low abundance, remains a significant bottleneck in proteoform measurement science. Manual validation by comparing observed fragments with intact masses and proteoform assignments is reliable yet time-consuming (e.g., Figure 3). The implementation of automated ranking (0–5) of proteoforms served to help gauge the quality of the fragmentation data for supporting the presence of one or more PTM and assisted with differentiating proteoform isobars. Other avenues for validation includes advanced scoring metrics such as the PTM characterization score (*C*-score⁵⁵), fraction collection, and reintroduction of specific proteoforms by direct infusion to improve fragment S/N, selective reaction monitoring assays of specific PTMs sites,⁵⁶ or wet lab procedures (e.g., immunoassays, cell culturing, etc.).

CONCLUSION

We present a DIA platform comprising SPLC-NSD-FTMS coupled with a three-tiered absolute mass search strategy that was employed for simultaneous protein/proteoform identifications from mouse brain extract. With a single 90-minute SPLC-NSD-FTMS run, the platform identified numerous proteoforms associated with four classical MBP ASVs including 323 unique proteoforms with a gradient optimized for the 14.1 kDa ASV. An offline three-tiered absolute mass search strategy enabled continuous proteoform identification and ranked validation across the LCMS run in an automated environment. The efficiency of the custom bioinformatics and data visualization workflow also enabled manual discovery of novel proteoforms that included two new splice variants with masses of ~12.1 and ~16.4 kDa that are similar to the 14.1 and 18.4 kDa splice variants with an ATSS. Similarly, we discovered S-acylation by fatty acid lipids (e.g., palmitic acid, stearic acid, and oleic acid) specifically modifying the 17.125 kDa splice variant. The robustness and speed of the procedures presented is expected to enable future characterization of MBP (or other highly modified proteins) in clinical proteomics investigations (e.g., MS disease models). Plus, the work contributes to the emerging field of proteoform measurement science, which complements protein-centric proteomics by tackling the breadth of proteoform space occupied by a single protein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

S.M.P. contributed to project concept, oversight, software/code, data tabulation, and writing. D.E.R. and C.E.W. contributed to software and data analysis. D.A.P. contributed to data analysis and writing. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1R01GM115739-01A1. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health. This work was also supported by the Multiple Sclerosis Society [PP-1503-04034], The Darrel K. Royal Research Fund for Alzheimer's Disease [48680-DKR], The Texas Alzheimer's Research and Care Consortium Investigator Grant Program [354091], and the UT System Neuroscience and Neurotechnology Research Institute [363027], and The John L. Roach Scholarship in Biomedical Research.

References

1. Boggs JM. Cell Mol Life Sci. 2006; 63:1945–1961. [PubMed: 16794783]
2. Schmitt S, Castelvetti LC, Simons M. Biochim Biophys Acta, Mol Cell Biol Lipids. 2015; 1851:999–1005.
3. Poliak S, Peles E. Nat Rev Neurosci. 2003; 4:968–980. [PubMed: 14682359]
4. Love S. J Clin Pathol. 2006; 59:1151–1159. [PubMed: 17071802]
5. Harauz G, Libich DS. Curr Protein Pept Sci. 2009; 10:196–215. [PubMed: 19519451]
6. Vassall KA, Bamm VV, Harauz G. Biochem J. 2015; 472:17–32. [PubMed: 26518750]
7. Harauz G, Boggs JM. J Neurochem. 2013; 125:334–361. [PubMed: 23398367]
8. Kim JK, Mastronardi FG, Wood DD, Lubman DM, Zand R, Moscarello MA. Mol Cell Proteomics. 2003; 2:453–462. [PubMed: 12832457]
9. Akiyama K, Ichinose S, Omori A, Sakurai Y, Asou H. J Neurosci Res. 2002; 68:19–28. [PubMed: 11933045]
10. Abbott WM, Mellor A, Edwards Y, Feizi T. Biochem J. 1989; 259:283–290. [PubMed: 2470348]
11. Harauz G, Ishiyama N, Hill CM, Bates IR, Libich DS, Fares C. Micron. 2004; 35:503–542. [PubMed: 15219899]
12. Pritzker LB, Joshi S, Harauz G, Moscarello MA. Biochemistry. 2000; 39:5382–5388. [PubMed: 10820009]
13. Zand R, Li MX, Jin X, Lubman D. Biochemistry. 1998; 37:2441–2449. [PubMed: 9485392]
14. Martenson RE, Law MJ, Deibler GE. J Biol Chem. 1983; 258:930–937. [PubMed: 6185481]
15. Zand R, Jin X, Kim J, Wall DB, Gould R, Lubman DM. Neurochem Res. 2001; 26:539–547. [PubMed: 11513482]
16. Kim J, Zhang R, Strittmatter EF, Smith RD, Zand R. Neurochem Res. 2009; 34:360–372. [PubMed: 18618245]
17. Zhang C, Walker AK, Zand R, Moscarello MA, Yan JM, Andrews PC. J Proteome Res. 2012; 11:4791–4802. [PubMed: 22420465]
18. Cao L, Sun D, Whitaker JN. J Neuroimmunol. 1998; 88:21–29. [PubMed: 9688320]
19. Constantinescu CS, Farooqi N, O'Brien K, Gran B. Br J Pharmacol. 2011; 164:1079–1106. [PubMed: 21371012]
20. Patrie SM. Adv Exp Med Biol. 2016; 919:171–200. [PubMed: 27975217]
21. Janssen KA, Sidoli S, Garcia BA. Methods Enzymol. 2017; 586:359–378. [PubMed: 28137571]
22. Zhang J, Corbett JR, Plymire DA, Greenberg BM, Patrie SM. Proteomics. 2014; 14:1223–1231. [PubMed: 24678018]
23. Zhang J, Roth MJ, Chang AN, Plymire DA, Corbett JR, Greenberg BM, Patrie SM. Anal Chem. 2013; 85:10377–10384. [PubMed: 24074297]

24. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li MX, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. *Nature*. 2011; 480:254–U141. [PubMed: 22037311]
25. Tian Z, Tolic N, Zhao R, Moore RJ, Hengel SM, Robinson EW, Stenoien DL, Wu S, Smith RD, Pasa-Tolic L. *Genome Biol* [Online]. 2012; 13doi: 10.1186/gb-2012-13-10-r86
26. Li M, Jiang L, Kelleher NL. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2009; 877:3885–3892.
27. Hanisch FG. *Methods Mol Biol* (N Y, NY, U S). 2012; 842:179–189.
28. Cai W, Tucholski TM, Gregorich ZR, Ge Y. *Expert Rev Proteomics*. 2016; 13:717–730. [PubMed: 27448560]
29. Gregorich ZR, Cai W, Lin Z, Chen AJ, Peng Y, Kohmoto T, Ge Y. *J Mol Cell Cardiol*. 2017; 107:13–21. [PubMed: 28427997]
30. Burnaevskiy N, Fox T, Plymire DA, Ertelt JM, Weigele BA, Selyunin AS, Way SS, Patrie SM, Alto NM. *Nature*. 2013; 496:106. [PubMed: 23535599]
31. Patrie SM, Charlebois JP, Whipple D, Kelleher NL, Hendrickson CL, Quinn JP, Marshall AG, Mukhopadhyay B. *J Am Soc Mass Spectrom*. 2004; 15:1099–1108. [PubMed: 15234368]
32. Patrie SM, Ferguson JT, Robinson DE, Whipple D, Rother M, Metcalf WW, Kelleher NL. *Mol Cell Proteomics*. 2006; 5:14–25. [PubMed: 16236702]
33. Maatta JA, Coffey ET, Hermonen JA, Salmi AA, Hinkkanen AE. *Biochem Biophys Res Commun*. 1997; 238:498–502. [PubMed: 9299539]
34. Pesavento JJ, Kim YB, Taylor GK, Kelleher NL. *J Am Chem Soc*. 2004; 126:3386–3387. [PubMed: 15025441]
35. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. *Nucleic Acids Res*. 2007; 35:W701–706. [PubMed: 17586823]
36. Wu S, Zhu Y. *Bioinformation*. 2012; 8:167–169. [PubMed: 22368391]
37. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. *Methods Mol Biol*. 2007; 406:89–112. [PubMed: 18287689]
38. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. *Nat Biotechnol*. 2001; 19:952–957. [PubMed: 11581661]
39. Roth MJ, Plymire DA, Chang AN, Kim J, Maresh EM, Larson SE, Patrie SM. *Anal Chem*. 2011; 83:9586–9592. [PubMed: 22017354]
40. Riley NM, Westphall MS, Coon JJ. *Anal Chem*. 2015; 87:7109–7116. [PubMed: 26067513]
41. Helbig AO, Gauci S, Raijmakers R, van Breukelen B, Slijper M, Mohammed S, Heck AJ. *Mol Cell Proteomics*. 2010; 9:928–939. [PubMed: 20061308]
42. Polevoda B, Sherman F. *J Mol Biol*. 2003; 325:595–622. [PubMed: 12507466]
43. Varland S, Osberg C, Arnesen T. *Proteomics*. 2015; 15:2385–2401. [PubMed: 25914051]
44. Ross NW, Braun PE. *J Neurosci Res*. 1988; 21:35–44. [PubMed: 2464071]
45. Destefano JJ, Schuster SA, Lawhorn JM, Kirkland JJ. *J Chromatogr A*. 2012; 1258:76–83. [PubMed: 22939204]
46. Kirkland JJ, Truszkowski FA, Dilks CH Jr, Engel GS. *J Chromatogr A*. 2000; 890:3–13. [PubMed: 10976789]
47. Schuster SA, Wagner BM, Boyes BE, Kirkland JJ. *J Chromatogr A*. 2013; 1315:118–126. [PubMed: 24094750]
48. Wagner BM, Schuster SA, Boyes BE, Shields TJ, Miles WL, Haynes MJ, Moran RE, Kirkland JJ, Schure MR. *J Chromatogr A*. 2017; 1489:75–85. [PubMed: 28213987]
49. Andjelkovic U, Tufegdizic S, Popovic M. *Electrophoresis* [Online]. 2017; Just accepted. doi: 10.1002/elps.201700260
50. Mohr J, Swart R, Samonig M, Bohm G, Huber CG. *Proteomics*. 2010; 10:3598–3609. [PubMed: 20859959]
51. Rogers BA, Wu Z, Wei B, Zhang X, Cao X, Alabi O, Wirth MJ. *Anal Chem*. 2015; 87:2520–2526. [PubMed: 25646567]
52. Wu Z, Wei B, Zhang X, Wirth MJ. *Anal Chem*. 2014; 86:1592–1598. [PubMed: 24383398]

53. Astefanei A, Dapic I, Camenzuli M. *Chromatographia*. 2017; 80:665–687. [PubMed: 28529348]
54. Bjelke B, Seiger A. *Int J Dev Neurosci*. 1989; 7:145–164. [PubMed: 2469297]
55. LeDuc RD, Fellers RT, Early BP, Greer JB, Thomas PM, Kelleher NL. *J Proteome Res*. 2014; 13:3231–3240. [PubMed: 24922115]
56. Kuzyk MA, Parker CE, Domanski D, Borchers CH. *Methods Mol Biol*. 2013; 1023:53–82. [PubMed: 23765619]

Figure 1.

(A) TIC chromatogram shows a typical elution profile of proteins obtained from Folch/acid extracts of homogenized mouse brain tissue. (B) Fragmentation-level feature map highlights the mass, retention time, and intensity (marker size) of fragments detected by continuous elution SPLC-NSD-FTMS processed at 1 min intervals. Fragments labeled highlight the methylated γ -ions for the 14.1 kDa MBP ASV and a non-MBP protein poly ubiquitin-b (U). (C) Protein-level feature map highlights elution profile of various MBP ASVs and non-MBP proteins identified within the sample presented in (A). Encircled are proteoform groups associated with the 12.1, 14.1, 16.4, 17.125, 17.140, and 18.4 kDa ASVs as well as various histone variants (e.g., histone 1.4 (H1.4), histone 1.2 (H1.2), histone H2B (H2B), etc.). Ratios of the classical ASVs determined from the summation of spectral intensity for their respective proteoform groups (inset).

Figure 2.
(A–D) Representative broadband mass spectra associated with the peak elution period (centroid) for the 14.1, 18.4, 17.125, and 17.140 kDa MBP ASVs, respectively (left). Insets (middle) shows the spectral complexity observed at the single charge state for the respective ASVs. The fragment map (right) for each respective ASV highlights the b-ions and y-ions for the N-ACET proteoform identified at the corresponding elution period from an SPLC-NSD-FTMS run.

Figure 3.

(A) Proteoform-level feature map associated with 14.1 kDa ASV. (B–G) Mass spectra that correspond to the respective regions labeled in (A). Labels a–m in the bar inset (A) designate apparent nominal mass proteoforms detected throughout the analysis. The bar insets (B–G) highlight common delta mass differences (e.g., 14, 16, 80 Da, etc.) observed between spectral partners for groups a–m across the elution period. Mass spectra insets of the b81 ion (left) highlight elution of a MSO-M19 proteoform early in the run (B–C vs D). Mass spectra insets of y43 ion (right) highlight the MSO-M124 proteoform prior to the METH-

R104 and DIMET-R104 later in the run (B–G). Ox: MSO; Met: METH; diMet: DIMETH; *: unmodified fragment.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

(A) Heat map highlights average P' over time of all unique proteoform identifications made for the apparent nominal mass groups shown in Figure 3A (see also Supplemental Table 2). (B) Expanded heat map highlights the relative contributions of the unique proteoforms (red) N-ACET (14 113 Da), METH (14 127 Da), MSO (14 129 Da), DIMETH (14 141 Da), METH+MSO (14 143 Da), and MSONE (14 145 Da) to their respective average P' (black). (C) Expanded heat map for the numerous unique PHOS (14 192 Da) and PHOS+METH (14 207 Da) proteoforms. Ac: N-ACET, Met: METH, Ox: MSO, dMet: DIMETH, dOx: MSONE, Phos: PHOS.

Figure 5.
(A,B) Mass spectra associated with a single charge state at the peak elution period (centroid) of the novel 12.1 and 16.4 kDa ASVs with fragment map confirming the identification.

Figure 6.

(A) Spectrum and fragment map confirmed the addition of palmitic acid (16:0) to the 17.125 kDa ASV through S-acylation of C78. (B) Several proteoforms associated with the addition of oleic acid (18:2), stearic acid (18:0), and other long chain fatty acids (*) were chromatographically resolved and detected by accurate intact mass only.