DEVELOPMENT OF A TOP-DOWN PROTEOMICS PLATFORM FOR CHARACTERIZING MYELIN BASIC PROTEIN PROTEOFORMS

APPROVED BY SUPERVISORY COMMITTEE

Steven Patrie, PhD

Yonghao Yu, PhD

Nancy Monson, PhD

Nick Grishin, PhD

DEDICATION

I thank Kristen Wilhite for encouragement and support.

I thank the Burkholder and Main families for friendship and support.

I thank my family for patience and support.

DEVELOPMENT OF A TOP-DOWN PROTEOMICS PLATFORM FOR CHARACTERIZING MYELIN BASIC PROTEIN PROTEOFORMS

by

DANIEL ALVAN PLYMIRE

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY IN MOLECULAR BIOPHYSICS
WITH A CONCENTRATION IN COMPUTATIONAL AND SYSTEMS BIOLOGY

The University of Texas Southwestern Medical Center

Dallas, Texas

August, 2017

DEVELOPMENT OF A TOP-DOWN PROTEOMICS PLATFORM FOR CHARACTERIZING MYELIN BASIC PROTEIN PROTEOFORMS

Daniel Alvan Plymire, MS

The University of Texas Southwestern Medical Center at Dallas, 2017

Supervising Professor: Steven M. Patrie, PhD

A top-down proteomic analysis platform for the analysis of myelin basic protein was designed and applied to murine central nervous system tissue. The compatibility of superficially porous (SP) resin for label-free intact protein analysis of myelin basic protein with on-line capillary liquid chromatography mass spectrometry was demonstrated and was able to resolve splice variants and individual proteoforms of myelin basic protein. Data independent acquisition using nozzle skimmer dissociation yielded fragment ions that were used with novel informatics procedures to identify proteoforms of MBP. The informatics procedures including continuous elution proteoform scoring, a three-tier search methodology, and a ranking system for identification validation yielded 323 unique proteoforms from a single splice variant of MBP from a single LC run. A novel site of acylation and two novel splice variants derived from a novel transcriptional start site were also identified.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

1. Proteoform analysis of lipocalin-type prostaglandin D-synthase from human cerebrospinal fluid by isoelectric focusing and superficially porous liquid chromatography with Fourier transform mass spectrometry. Zhang J, Corbett JR, Plymire DA, Greenberg BM, Patrie SM. Proteomics. 2014 May;14(10):1223-31. doi: 10.1002/pmic.201300368. Epub 2014 Mar 20. PMID: 24678018

2. Measurement of blood protease kinetic parameters with self-assembled monolayer ligand binding assays and label-free MALDI-TOF MS. Patrie SM, Roth MJ, Plymire DA, Maresh E, Zhang J. Anal Chem. 2013 Nov 5;85(21):10597-604. doi: 10.1021/ac402739z. Epub 2013 Oct 23. PMID: 24107006

3.Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography. Zhang J, Roth MJ, Chang AN, Plymire DA, Corbett JR, Greenberg BM, Patrie SM. Anal Chem. 2013 Nov 5;85(21):10377-84. doi: 10.1021/ac402394w. Epub 2013 Oct 16. PMID: 24074297

4. Proteolytic elimination of N-myristoyl modifications by the Shigella virulence factor IpaJ. Burnaevskiy N, Fox TG, Plymire DA, Ertelt JM, Weigele BA, Selyunin AS, Way SS, Patrie SM, Alto NM. Nature. 2013 Apr 4;496(7443):106-9. doi: 10.1038/nature12004. Epub 2013 Mar 27. PMID: 23535599

5. Surface preparation strategies for improved parallelization and reproducible MALDI-TOF MS ligand binding assays. Roth MJ, Maresh EM, Plymire DA, Zhang J, Corbett JR, Robbins R, Patrie SM. ACS Appl Mater Interfaces. 2013 Jan;5(1):6-10. doi: 10.1021/am3024756. Epub 2012 Dec 24. PMID: 23249094

6. MS-based ligand binding assays with speed, sensitivity, and specificity. Roth MJ, Kim J, Maresh EM, Plymire DA, Corbett JR, Zhang J, Patrie SM. Proteomics. 2012 Nov;12(21):3143-6. doi: 10.1002/pmic.201200151. Epub 2012 Oct 12. PMID: 22965683

7. Thin-layer matrix sublimation with vapor-sorption induced co-crystallization for sensitive and reproducible SAMDI-TOF MS analysis of protein biosensors. Roth MJ, Kim J, Maresh EM, Plymire DA, Corbett JR, Zhang J, Patrie SM. J Am Soc Mass Spectrom. 2012 Oct;23(10):1661-9. Epub 2012 Jul 31. PMID: 22847391

8. Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry. Roth MJ, Plymire DA, Chang AN, Kim J, Maresh EM, Larson SE, Patrie SM. Anal Chem. 2011 Dec 15;83(24):9586-92. doi: 10.1021/ac202339x. Epub 2011 Nov 15. PMID: 22017354

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

LIST OF TABLES

LIST OF APPENDICES

LIST OF DEFINITIONS

ATSS	Alternative Transcriptional Start Site
EAE	Experimental Autoimmune Encephalomyelitis
ESI	Electrospray Ionization
FTMS	Fourier Transform Mass Spectrometry
GRAVY	Grand Average of Hydropathy
ISD	In Source Dissociation
LC	Liquid Chromatography
MBP	Myelin Basic Protein
NSD	Nozzle Skimmer Dissociation
PAGE 1D	Polyacrylamide Gel Electrophoresis
PTM	Post-Translational Modification
RCF	Relative Centripetal Force
RPLC	Reversed Phase Liquid Chromatography
S/N	Signal-to-Noise Ratio
SP	Superficially Porous
SPLC	Superficially Porous Reversed Phase Liquid Chromatography

CHAPTER ONE

INTRODUCTION TO TOP-DOWN PROTEOMICS AND MYELIN BASIC PROTEIN

Top-Down Proteomics

The underlying goal of top-down proteomics is to study all of the proteoforms of a given gene, which encompass all of the different protein species that result downstream from a given gene, including single or multiple nucleotide polymorphisms, alternative splicing, and post-translational modification [1, 2]. The top-down philosophy employs fragmentation strategies to intact proteins, enabling the discovery concomitant events that result in different proteoforms derived from a given gene. This is conceptually illustrated in Figure 1, where a single gene gives rise to multiple proteoforms that result from alternative splicing, nucleotide polymorphisms, and post-translational modifications. Figure 2 illustrates the exponential expansion of the set of proteoforms derived from a given gene product. The number of proteoforms in the proteome is given by Nps = N * ASav * SAPav * PTMav. (Nps: number of protein species; N: total number of protein encoding genes; ASav: average number of splice variants per protein encoding gene; SAPav: average number of nonsynonymous single nucleotide polymorphisms; PTMav: average number of PTM events per protein encoding gene) [3]. This equation can be modified for each splice variant of MBP as $\mathrm{Nsv}=\prod_{i=1}^{n} St_i$. (Nsv: the total number of proteoforms per splice variant; n: the number of amino acids that may harbor a PTM; St: the total number of modification states possible at each possible PTM site *i*.

The evolution of the top-down philosophy has encompassed protein separations, high resolving power mass spectrometry and informatics / data processing procedures [4]. The top-down philosophy has been employed in studies using proteoform level information to characterize histones [5], cardiac troponin I [6] and membrane transporters [7] as well as other targets. Reversed phase protein separations have employed a number of different strategies, including monolithic [8, 9], polymeric [10], porous silica [11], and nonporous silica [12]. Alternatives to reversed phase separations include hydrophobic interaction liquid chromatography (HILIC) in which the stationary phase is polar and the mobile phase increases in the proportion of aqueous phase [13, 14], and ion exchange chromatography in which separations rely on differential charges on the analyte [11].

My work evaluated superficially porous reversed phase liquid chromatography (SPLC) for intact protein separations, and found that it was compatible for intact protein analysis both with protein standards, complex mixtures [15], and myelin basic protein. Orthogonal separations such as IEF [16] and GELFrEE [17] have also been employed to prefractionate intact proteins prior to LC analysis. IEF fractionation is dependent on the isoelectric point of proteins while GELFrEE fractionation is dependent on molecular weight [17, 18]. Top-Down procedures may employ multiple separations using both on-line and off-line separations to achieve separation of complex mixtures [4].

A number of fragmentation strategies have been employed in top-down proteomics including electron transfer dissociation (ETD) [19], electron capture dissociation (ECD) [20],

ultraviolet photodissociation (UVPD) [21], higher-energy collisional dissociation (HCD) [21], and collision induced dissociation (CID) [18]. Advantages to ETD and ECD include improved coverage of proteins with labile modifications (i.e., phosphorylation) [22]. Electron based dissociation also tends to yield more c-ions and z-ions while the collisional methods yield more b-ions and y-ions [4]. The fragmentation strategies above all employ a data-dependent methodology where the mass spectrometer first generates a full-scan mass spectra of the species present and then acquires targeted MS spectra in a second stage of analysis on a set of those precursor species [23]. An alternative method uses heat-assisted nozzle-skimmer collisionally induced dissociation. This methodology is data-independent and all of the intact proteins entering the instrument are susceptible to gas-phase fragmentation [24, 25]. An advantage of data-independent acquisition is that fragmentation data may be collected on all of the species entering the mass spectrometer simultaneously. Mass spectrometry in top-down studies typically employs either Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) or Orbitrap mass spectrometry [4]. FT-ICR MS has employed CID and ECD and nozzle-skimmer dissociation (NSD) for fragmentation studies [26-29]. Orbitrap MS has employed NSD, CID, HCD, and ETD [30-33] for fragmentation studies.

A number of software packages have been developed to aid in the interpretation of top-down data. ProSight PTM employs absolute mass searching for the identification of intact proteins [34, 35]. The absolute mass search uses the intact mass to select the database targets that the experimental fragmentation data will be searched against. The absolute match search allows

the user to input a tolerance on both the accuracy of the intact mass and the accuracy of the fragments that were observed experimentally. ProSight PC delivers a p-score and e-score used to characterize the confidence in the identification by accounting for the probability that a random sequence could provide an equally supportive number of fragment matches [36]. The p-score is dependent on the number of fragments searched against, the largest matching fragment, and the number of fragment matches. The e-score multiplies the p-score by the database size to better characterize the probability of false positives in the database. ProSight PTM 2.0 includes the ability to add fixed modifications and terminal modifications to the search functionality of ProSight PTM [37]. The Δm search mode in ProSightPC 2.0 enables discovery of unknown PTMs. This search mode takes the difference between the observed and theoretical precursor masses and performs two additional queries including the difference between the observed and theoretical precursor masses being added and subtracted from the observed fragment list. This search mode is useful for the characterization of proteoforms where there are unknown modifications that are not reported in a database such as UniProt. Data preprocessing to generate precursor and fragment masses from mass spectra may employ algorithms such as Thorough High Resolution Analysis of Spectra by Horn (THRASH) [38] and Xtract [39]. These methods identify precursor and fragment masses from potential isotopic clusters in each spectrum.

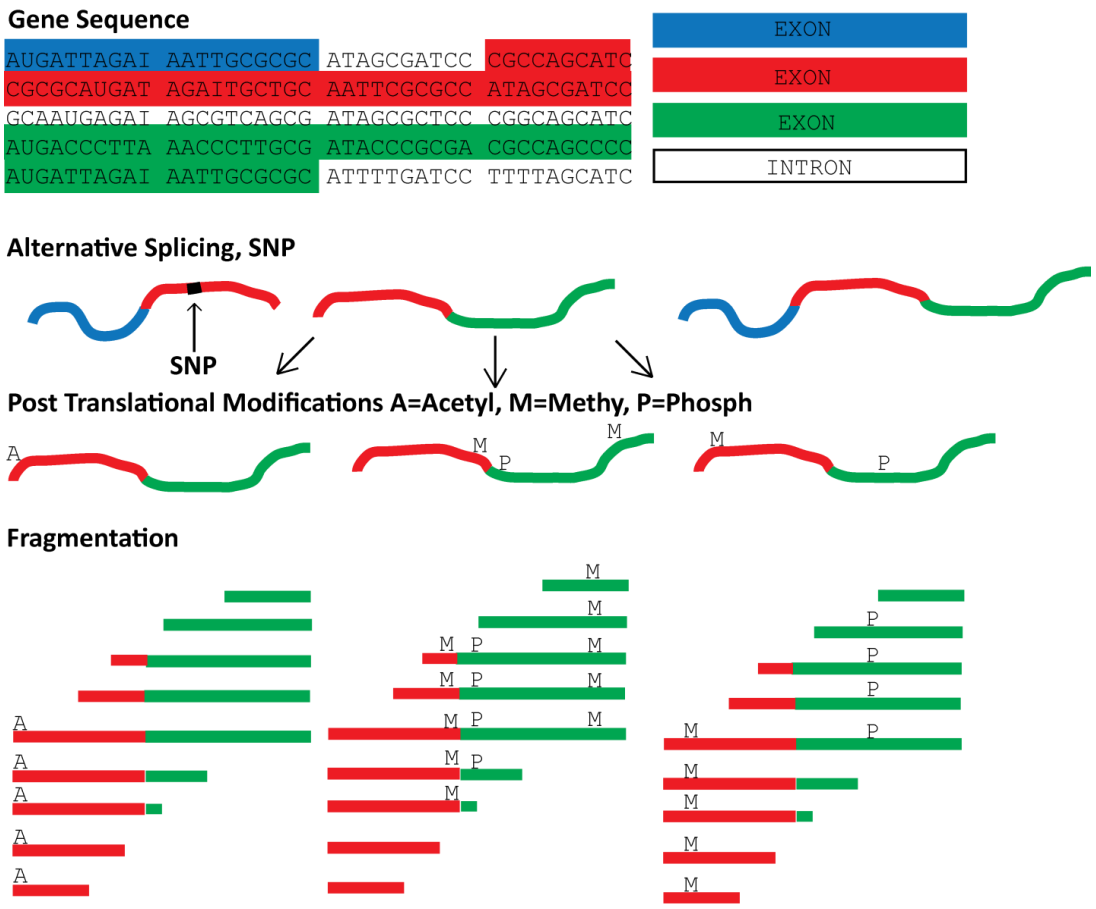


Figure 1. The top-down proteomics philosophy. A given gene sequence may have a number of exons that can be combined in different combinations during alternative splicing. These alternative splice variants may harbor SNPs or other nucleotide polymorphisms. Each splice variant may harbor different combinations of PTMs. Intact protein fragmentation generates fragments that can be used to unambiguously localize concomitant post-translational modifications.

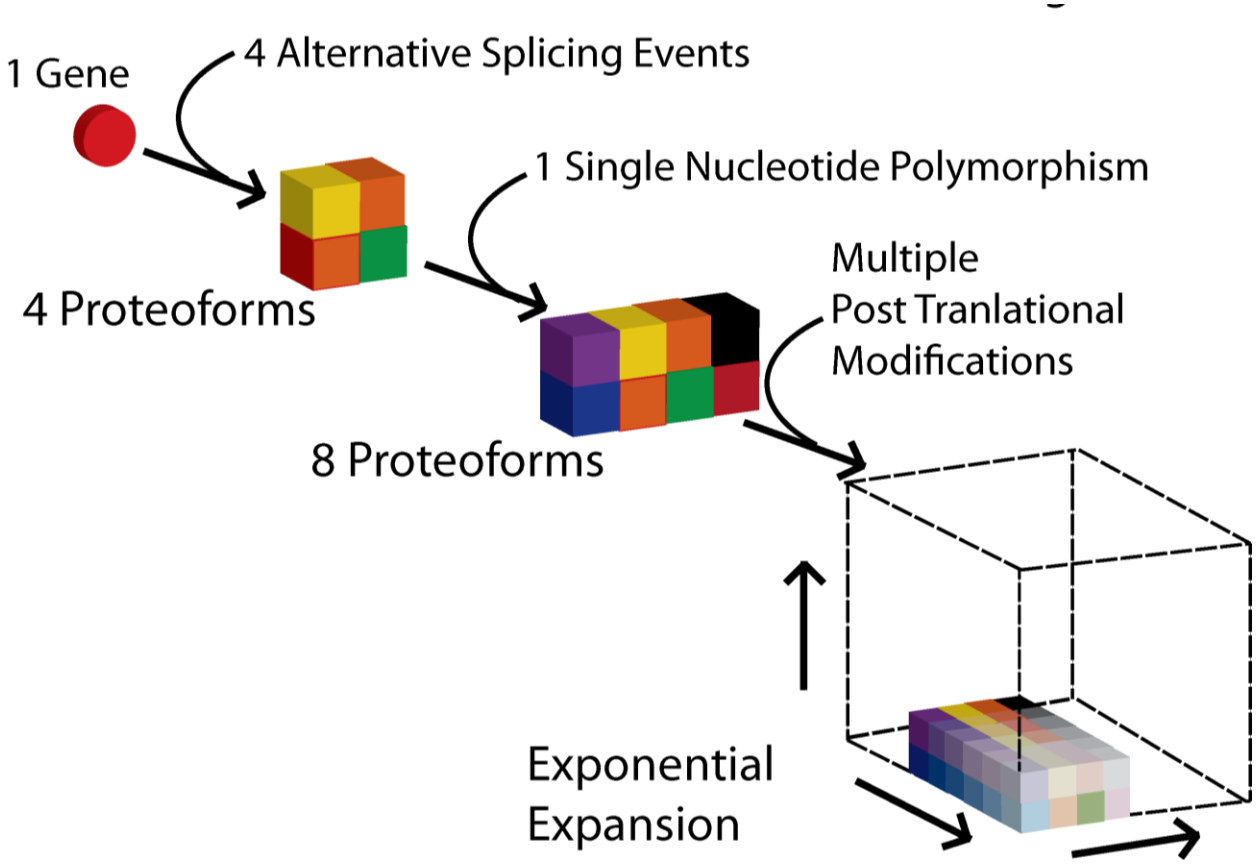


Figure 2. Increases in the number of proteoforms that arise from a single gene. A single gene can give rise to multiple proteoforms resulting from alternative splicing, single nucleotide polymorphisms, and post-translational modification, resulting in a large set of potential proteoforms that could be observed with top-down proteomics.

Introduction to Myelin Basic Protein

Myelin basic protein (MBP) was described as an "executive" molecule by Mario Moscarello and the proteoform complexity related to splice variants of MBP and the influence of PTMs in the myelin sheath is analogous to the influence of PTMs in the epigenetic code of histones in chromatin [40]. MBP plays an essential role in myelin compaction, and post-translational

modification of MBP has been implicated in the progression of multiple sclerosis in humans [40].

The myelin sheath consists of multiple layers and is produced by oligodendrocytes in the central nervous system (CNS) [41, 42]. The myelin sheath serves to aid in the propagation of signal transduction down the axon and the integrity of the myelin sheath determines the speed of saltatory signal transduction in the nerve [43]. Myelin basic protein is a major constituent of the myelin sheath and makes up approximately 30% of the dry protein mass of the myelin sheath [41]. Disruption of the myelin sheath can lead to neurodegeneration in disorders such as multiple sclerosis.

MBP is known to harbor a myriad of PTMs, including N-terminal acylation, serine and threonine phosphorylation, arginine mono-methylation and di-methylation, arginine deimination (citrullination), asparagine deamidation, and ADP ribosylation [44]. When including post-translational modification sites based on similarity with all other organism, a total of over 60 PTMs have been reported on the 18.4 kDa splice variant. Alternative splicing results in a number of different variants from the 10 exons in human and the 11 exons in murine tissue [42]. The presence of multiple splice variants of MBP and extensive post-translational modification of MBP makes MBP highly heterogeneous and the top-down philosophy is the only method that can identify all sources of heterogeneity that are present.

Multiple Sclerosis

Epidemiology and Costs

Multiple sclerosis is a neurodegenerative disease of the central nervous system that typically presents in the third and fourth decade of life [45]. In young adults, multiple sclerosis is the second leading cause of acquired disability after trauma [46]. Caucasians may harbor a genetic predisposition to multiple sclerosis and there are distinct geographic and ethnic distributions of multiple sclerosis susceptibility [47]. The clinical disease course and disease severity of multiple sclerosis also varies by ethnic group [48].

Total direct and indirect costs of multiple sclerosis are estimated at $47,215 per patient year (2004 US dollars) with disease modifying drugs make up much of the direct costs [49]. The estimated costs of multiple sclerosis total $14 billion (2004 US dollars) per year in the United States [50]. Indirect costs are significant with multiple sclerosis, as 50% of patients reach loss of employment, require use of assisted walking devices, and acquire the inability to walk in 10, 15, and 25 years, respectively [51].

Etiology

Multiple sclerosis is thought to be an autoimmune disease mediated by $CD4^+$ T cells [52]. Multiple sclerosis is polygenic and the only known region associated with multiple sclerosis is the MHC region on chromosome 6 [53]. Approximately 15-20% of multiple sclerosis patients have a family history of multiple sclerosis although there is no clear mode of

inheritance [53]. Postgermline events may influence the disease course of multiple sclerosis [48].

Diagnosis of Multiple Sclerosis

There are several subtypes of multiple sclerosis including relapsing-remitting (relapses with periods without clinical progression), secondary progressive (gradual deterioration following relapsing-remitting disease), primary progressive (disease progression from the onset) and progressive-relapsing (acute exacerbations with a primary progressive disease course) [45]. Diagnosis was once based exclusively on clinical history and physical examination [47]. Diagnosis of multiple sclerosis requires "exclusion of multiple sclerosis mimics, diagnosis of isolated clinical syndromes and differentiating between multiple sclerosis and non-multiple sclerosis inflammatory demyelinating diseases [54]."

Noseworthy et al., noted many clinical symptoms characteristic of multiple sclerosis [55]. It was noted that "relapsing remitting multiple sclerosis starts with sensory disturbances, unilateral optic neuritis, diplopia, limb weakness, clumsiness, gait ataxia, bladder and bowel symptoms," and "eventually cognitive impairment depression, emotional lability, dysarthria, dysphagia, vertigo, ataxic tremors, pain sexual dysfunction spasticity become troublesome" [55].

The first episode of inflammatory demyelination is known as clinically isolated syndrome [45]. The Poser criteria defined diagnosis of multiple sclerosis by two or more demyelinating

attacks with two or more parts of the CNS involved [47]. The later McDonald diagnostic criteria [56] include evidence from MRI data and altered the Poser criteria by allowing new lesions to make up the second demyelinating event [45].

The McDonald criteria were introduced in 2001, modified in 2005 [57], and modified again in 2010 [58]. The McDonald criteria utilize data from magnetic resonance imaging and may diagnose patients presenting with a clinically isolated syndrome and serves have a greater prognostic role than other diagnosis criteria [54, 59]. Revisions to McDonald criteria improved specificity and sensitivity and allow for more rapid diagnosis [58]. McDonald criteria include both clinical presentation and additional data to substantiate multiple sclerosis diagnosis. The additional data include dissemination in space, dissemination in time, clinical attacks, positive CSF, or continued progression [56, 60].

In terms of disease progression and prognosis, most multiple sclerosis patients have a relapsing-remitting disease course [61]. The clinical course of multiple sclerosis is very heterogeneous, but secondary progression typically leads to irreversible, progressive disability [61]. Relapses may develop over hours to days and persist for days to weeks before dissipating [53]. Increases attack frequency or poor recovery suggest a poor prognosis for future degeneration [53]. Thirty percent of patients diagnosed with a clinically isolated syndrome progress to multiple sclerosis within 12 months [62]. Over half of patients make a transition to secondary progressive multiple sclerosis within 10 to 20 years after diagnosis

[50]. Life expectancy has been reported to be reduced by 4 to 12 years [50]. 10 to 20% of multiple sclerosis patients remain unimpaired without therapy for 20 years [51].

Mechanism of Multiple Sclerosis

The four key pathological features of multiple sclerosis include inflammation, demyelination, axonal loss and gliosis. The major responses to these pathological features include inflammation and remyelination [45].

Multiple sclerosis may be the result of an autoreactive T-cell-mediated cellular immune response against myelin [63, 64]. The mechanism of the pathogenesis of MBP is thought to involve T cells in the periphery becoming activated. These T cells may produce inflammatory cytokines and may differentiate into Th1 or Th17 cells. Activated T cells may cross the blood brain barrier, and release of cytokines by T cells may attract B cells and macrophages to the central nervous system. T cell contact with a myelin antigen results in reactivation and differentiation leading to a further inflammatory response. Activated macrophages that have been recruited may than damage the central nervous system. Antibodies may also contribute to the inflammatory demyelination and exists in plaque tissues [45, 53] . Elements of the humoral immune response may also contribute. The result is the destruction of myelin and eventual degeneration of axons [45]. Systemic infection could also cause up-regulation of adhesion molecules on the endothelium in the brain and spine that enable leukocytes to cross vessel walls [65].

The myelin sheath is the primary target of damage in multiple sclerosis, but axons, nerve cells and astrocytes are also affected [59]. Axonal degeneration is thought to contribute to permanent neurological deficits in multiple sclerosis and "axonal loss is variable with axonal density within plaques ranging from 20% to 80% of that in healthy white matter [48]." Inflammatory cytokines may inhibit axonal function [55]. Deficits from inflammation and remyelination are reversible, while axonal degeneration is thought to be permanent [48]. Loss of neurologic function is the result of axonal injury, gliotic scarring, and exhaustion of the oligodendrocyte progenitor pool [55].

Treatments for Multiple Sclerosis

The first line of treatment for multiple sclerosis includes type 1 interferons (IFN) and glatiramer acetate [45]. Interferon-β inhibits T-cell activation reduces the permeability of the blood brain barrier to inflammatory cells [61]. Glatiramer acetate is a synthetic co-polymer similar to MBP that functions by inhibition of MBP-reactive t-lymphocytes and induction of T helper lymphocytes in the central nervous system [61]. The second line of treatment includes natalizumab [45]. Natalizumab I a monoclonal antibody that prevents transmigration of inflammatory lymphocytes across the blood brain barrier [61]. Another treatment, mitoxantrone is cytotoxic and immunosuppressive and may inhibit T cell activation, suppress B cell, T cell, and macrophage proliferation, impair antigen presentation, prevent demyelination and reduce pro-inflammatory cytokines [61]. Therapies may reduce the frequency of relapse and delay disease progression in relapsing-remitting and secondary progressive multiple sclerosis [66].

Experimental Autoimmune Encephalomyelitis

Background to Experimental Autoimmune Encephalomyelitis

Induction of EAE in different species including macaque, guinea pig, rabbit, mouse, rat, hamster, dog, sheep and marmoset have different clinical signs and lead to different pathological findings [67]. The response to EAE induction is dependent on the mode of sensitization, the immunogen, and the organism's genetic background [68]. EAE has been induced by whole myelin homogenate, MBP, myelin oligodendrocyte glycoprotein (MOG), and proteolipid protein (PLP) [68]. MBP and peptides derived from MBP are encephalitogenic an induce EAE in rodents and primates [69]. MBP-specific CD4+ T cells that are primed in vitro can cause EAE upon introduction to normal or immunocompromised mice [70]. Epitope spreading has been noted in the EAE model, whereby injection of a single myelin protein epitope leads to activation against other epitopes of the same protein [53].

Constantinescu et al., enumerated uses of EAE including study of "neuroprotective strategies, immunosuppressive drugs, neurotransmitters in inflammation, channel function during inflammation, demyelination, and remyelination, immune responses in immunologically privileged sites, effects of cytokines on the central nervous system, blood-brain barrier function and dysfunction, immunological tolerance, T-cell receptor restriction, epitope spreading, and regulatory T cells" [45]. EAE was also used in development of glatiramer acetate, mitoxantrone, and natalizumab [66]. Some EAE models reflect aspect of multiple sclerosis including axonal degeneration and demyelination [66]. The disease course of EAE

may be monophasic, chronic relapsing, or primary progressive [68]. As an example, the acute monophasic disease course can be induced in SJL mice using spinal cord homogenate, the relapsing-remitting disease course can be induced in SJL mice by PLP139-151, and the chronic disease course can be induced in C57BL/6 mice with MOG35-55 [68]. Relapsing experimental autoimmune encephalomyelitis is $CD4^+$ T cell-mediated initiated by immunizing SJL, pl/J or B10.PL mice with myelin antigens [64].

Myelin

Myelination of axons serves to ensure saltatory action potential transmission in vertebrates [71]. 70% of myelin (dry weight) is made of lipids and the major protein components are proteolipid protein and myelin basic protein [44]. Myelin in patients with multiple sclerosis is structurally unstable, and changes in MBP expression including both post-translational modification and alternative splicing may facilitate the changes in myelin structure that contribute to demyelination in multiple sclerosis [69]. An understanding of the pathogenesis of neurodegenerative disease requires understanding of the mechanism of changes in the myelin sheath.

Lesions

Lesions are demyelinated plaques consisting of "well-demarcated hypocellular area characterized by loss of myelin, preservation of axons, and formation of astrocyte scars" [55]. Lesions are most common in the optic nerves periventricular white matter, brain stem, cerebellum, and spinal cord white matter [55]. The location of lesions may influence the

extent of their remyelination as it was shown that cortical lesions were often well remyelinated, whereas those in the peri-ventricular regions were not [46]. Lesions may have diagnostic potential as the number of lesions on T2-weighted MRI, the number of lesions on contrast-enhanced MRI, and the volume of lesions of MRI are prognostic [62]. Additionally, examination of lesions from several patients with rapidly deteriorating relapsing and remitting multiple sclerosis showed extensive oligodendrocyte apoptosis and few infiltrating lymphocytes or phagocytes [63].

Three major types of MS lesions include acute plaques, chronic active plaques and chronic plagues [72]. These sites were characterized by Lock et al.: acute MS lesions included “areas of recent inflammation and edema, ongoing demyelination, abundant astroglial hypertrophy, and an ill-defined margin showing myelin vacuolation but little or no fibrous astrologists.” Chronic active lesions included “areas of long-term demyelination and fibrous astrologic the well-demarcated margins of which had superimposed regions of recent inflammation and ongoing demyelination”, and chronic silent lesions which have “no inflammatory activity, abundant gliosis, and well-demarcated margins” [73].

Others have characterized acute lesions by infiltration of immune cells and demyelination [45], active demyelination in with myelin phagocytes interacting with infiltrating T cells [63], or infiltration by macrophages and activated microglial cells [59]. Chronic lesions have also been characterized as the result of loss of myelin with few immune cells infiltrating and include the presence of gliosis [45].

Demyelination

In terms of disease, relapse in multiple sclerosis the result of acute inflammatory demyelination, and progression in multiple sclerosis is the result of chronic demyelination, gliosis, and axonal loss [48]. Immunohistochemistry studies suggested that acute axonal injury occurs during early and late active demyelination [74]. Four distinct patterns of demyelination in multiple sclerosis have been proposed, facilitated by: 1. macrophage toxins, 2. autoantibodies, 3. loss of MAG myelin associated glycoprotein and 4. degeneration of oligodendrocytes [75]. The role of oligodendrocyte populations in demyelination is unclear, some lesions have oligodendrocyte depopulation, while others have oligodendrocytes or oligodendrocyte progenitors present [46].

Remyelination

Prompt remyelination may help to prevent axonal degeneration [46] and regression of symptoms resulting from resolution of inflammation and partial remyelination [55]. Spatially, remyelination is typically present in early lesions as well as in satellite regions of large lesions [65]. Strategies for promoting remyelination include cell transplantation, growth factor therapy and antibody therapy [76]. During multiple sclerosis, remyelination is variable, with 20% of multiple sclerosis patients examined at autopsy had 60-96% remyelination, while 67% of multiple sclerosis patients had less than 25% remyelination [46].

Myelin Basic Protein

MBP Background

MBP is highly heterogeneous due to alternative splicing and post-translational modification. The role of MBP in the central nervous system is to maintain and compact the myelin sheath by bringing apposing faces of the oligodendrocyte membrane together [69]. MBP fulfills its structural role because of its positive charge and synergistic protein-lipid interactions [69]. Different proteoforms of MBP may be involved in signal transduction pathways in myelination and development, and potentially remyelination attempts in multiple sclerosis [44]. MBP has also been associated with molecules of the immune system [44]. There is some evidence of release of MBP in multiple sclerosis, as examination of CSF in multiple sclerosis patients found that 52 out of 72 patients had higher MBP content than the control group [77].

Charge Isomers

Early work separating the proteoforms of MBP was carried out by Moscarello and colleagues and relied on charge based separations on a CM-52 cation exchange column. MBP was separated into charge isomers C1-C8, with C1 being the least modified and most positive [44]. The majority of the change in charge was the result of deimination reducing the net positive charge of MBP. C8 myelin is more abundant in immature myelin and myelin from multiple sclerosis patients, and MBP from multiple sclerosis tissue was less cationic relative to normal tissue [69, 78]. A recombinantly generated charge isomer of MBP "C8" was found to have less membrane depth penetration and shorter α-helical structure. These structural

changes lead to MBP being more exposed to the cytosolic space where it may be digested by proteases [69]. Functionally, changes in the charge of MBP may facilitate a regulatory mechanism or degradative mechanism in multiple sclerosis [69].

Post-Translational Modification

Changes in PTM expression have been associated with disease as mono- and di-methylated arginine was increased in multiple sclerosis patients, deimination was elevated in multiple sclerosis patients, and phosphorylation was reduced in multiple sclerosis patients [78]. The biophysical changes associated with these modifications may influence the structure of MBP and its biological interactions. Methylation of MBP is irreversible and stabilizes MBP's structure and reduces its vulnerability to proteolysis [44]. Methylation is associated with compact myelin, and methylation during multiple sclerosis may reflect attempts at remyelination [44, 78]. Deimination also contributes to increasing MBP exposure in protein-lipid complexes leading to digestion by cathepsin D, and overall MBP may become more unstructured [44]. Deimination may both impede MBP's ability to fulfill its structural role and also expose an immunodominant epitope to proteases [69, 78]. Deimination of MBP is catalyzed by peptidylarginine deiminases (PADs) including PAD1, PAD2, PAD3, PAD4/PAD5, and PAD6 [42]. Upregulation of PAD may account for the higher ratio of deiminated MBP in both early development and multiple sclerosis [44]. Increased PAD is correlated to more severe multiple sclerosis [44], and PAD2 and PAD4 enzymes are increased in the brain tissue of multiple sclerosis patients [79]. Upregulation of PAD2 in normal appearing white matter may be the result of methylation of CpG islands in the PAD2I

gene promotor [80]. In multiple sclerosis deimination reduces the formation of β-structure, leading to protease degradation [78]. Deimination of MBP also enhances responsiveness to T cells, alters conformation, and induces vesicle fragmentation *in vitro* [78]. The changes in MBP's *in vivo* function after deimination may be a result of both reduction of net charge and changes in ligand binding [44]. Deimination is thought to be irreversible [42], and in patients with multiple sclerosis, the degree of deimination of arginine has been associated with disease progression [57, 60]. Individuals with chronic MS have approximately 50% of their MBP molecules deiminated, while only 20% of MBP molecules are deiminated in normal brain. In the case of Marburg's syndrome, an acute case of multiple sclerosis, over 90% of arginine residues are deiminated [42].

Phosphorylation alters the structure of MBP by increasing the amount of the protein in β-strand conformation and renders MBP to be less vulnerable to proteolysis [44]. Phosphorylation by protein kinase C in vitro increased the proportion of the protein in β-structure, particularly when S7 was phosphorylated due to interactions with R5 and R9 [78].

Comparison of MBP in myelin between children and multiple sclerosis patients showed that myelin in multiple sclerosis is developmentally immature. MBP from myelin of multiple sclerosis patient was found to be at the maturity of the MBP of a child less than six-year old [81].

Post-translational modification also has several biophysical consequences as including that the ability to bind and aggregate lipid vesicles is reduced in MBP after deimination [44]. Also, deimination of MBP by peptidyl arginine deiminase was 2x greater when MBP was nonmethylated suggesting a temporal order of modifications may be present [78].

The phosphorylation of MBP is expected as phosphorylation is common on intrinsically disordered proteins [44]. Major phosphorylation sites of MBP correspond with disordered regions of the protein, and phosphorylation of exon II containing splice variants is required for nuclear translocation [44].

MBP Biophysics

MBP has a high net positive charge (p*I* > 10) *in vivo*, and MBP is known to be intrinsically disordered and without a solid secondary or tertiary structure, with approximately 75% of its residues in a random coil conformation and some core elements of α-helix and β-sheet [42, 44, 82]. MBP binds divalent cations including Hg^{2+}, Cu^{2+}, Zn^{2+}, Mg^{2+}, Cd^{2+} and Co^{2+} which may influence its structure and interactions with the phospholipid bilayer [44].

Typically, the structure of intrinsically disordered proteins is dependent on their environment or the presence of a ligand [44]. Intrinsically disordered proteins typically have more P, E, K, S, and Q residues, and fewer W, Y, F, C, I, L, and N residues. While MBP is intrinsically disordered in solution, the presence of lipids or detergents enables it to gain α-helix and β-sheet character [44]. The primary amino acid sequence of MBP has enriched proline content,

which is associated with open conformations. MBP contains higher levels or arginine and lower levels of glutamic acid which helps to cause MBP to have a positive charge which serves to facilitate MBP's interaction with the negatively charged phospholipid bilayer and contributes of MBP's positive charge *in vivo* [42]. MBP also has depletion of isoleucine, leucine, tryptophan, and tyrosine residues, which is also characteristic of many intrinsically disordered proteins [42].

Solving the structure of MBP as well as the effects of deimination will yield a better understanding of the process of demyelinating disease [44]. Techniques for studying the structure-function of MBP include X-ray diffraction, transmission electron microscopy, NMR spectrometry of proteins or protein/lipid complexes, NMR spectrometry of solid-state lipid-protein assemblies [44]. X-ray diffraction may be challenging because of the challenges of membrane protein crystallization and co-crystallization [44]. Solving the 3D structure including the structural effects of deimination will yield an understanding of myelin architecture and the mechanisms of demyelinating disease.

Splice Variants

The dominant splice variant in humans is the 18.4 kDa splice variant, and in rodents it is the 14.1 kDa splice variant [44]. Splice variants containing exon II are not membrane associated and may be localized to the nucleus [44]. Expression of splice variants with exon II is increased during myelination and remyelination processes [44]. The splice variants without

exon II are typically membrane associated [44]. Golli-MBPs may be involved with signal transduction or may help to regulate gene transcription [44].

Biomarkers

No FDA approved surrogate marker (clinical, radiologic, immunologic) exist that can predict the transition from clinically isolated syndrome to multiple sclerosis [55, 62] or worsening multiple sclerosis [66]. It would be beneficial to be able to subtype therapeutic options based on biomarkers [75]. Berger et al., reported that antibodies to MBP are present in early multiple sclerosis and antibodies to MOG cause demyelination in vitro and in animal models of multiple sclerosis and have been found in lesion of patients with multiple sclerosis [62]. Berger at al., demonstrated that multiple sclerosis patients with anti-MBP and anti-MOG antibodies had a poorer prognosis with more relapses than multiple sclerosis patients without those antibodies and earlier relapses than patients without those antibodies [62]. It was also shown that anti-MOG and anti-MBP antibodies were associated with increased risk of relapse, but the number of lesions on MRI was not [62]. This results in the consequence that patients that are not seropositive for anti-MOG and anti-MBP antibodies may benefit from delaying immunomodulatory therapy [62]. The use of autoantibodies to MBP and MOG to diagnosis patients with clinically isolated syndrome is inexpensive and precise and may aid in patient counseling and treatment [62].

Other potential biomarkers relevant to multiple sclerosis diagnosis include neuromyelitis optica-IgG to identify patients with neuromyelitis optica versus multiple sclerosis [54]. In

multiple scleroiss, beta-amyloid-precursor-protein may be a marker of acute axonal injury [74]. Higher resolution magnetic resonance imaging can be used to examine changes in normal appearing white matter [54]. There is an increase in IgG in the CSF of patients with multiple sclerosis [53]. DNA demethylase has also been increased in normal appearing white matter in multiple sclerosis patients [83], and may antagonize the effects of deimination [84]. There has also been evidence of increased deimination of histone H3 in multiple sclerosis patients [85]. Overall, diagnostic a prognostic biomarkers can be of benefit when patients at high risk could be offered disease-modifying treatments [62].

Future Directions

A number of outstanding questions still are unanswered with respect go myelin basic protein and its role in multiple sclerosis pathology. The immunological and neurobiological mechanisms that give rise to the different disease courses of multiple sclerosis are not fully elucidated, and the role of MBP in those processes is not fully understood [59]. Discovery of clinical biomarkers to help define disease subtypes is inadequate, and expression of MBP as various proteoforms may serve as a biomarkers [59]. This characterization could potentially lead to subtype-specific therapy [59]. Study of MBP may also help to rationalize if recent therapies targeting specific mechanisms of demyelination or axonal injury will be relevant in multiple sclerosis patients [59].

References

1. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
2. Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 499-519.
3. Ponomarenko, E.A., et al., *The Size of the Human Proteome: The Width and Depth.* Int J Anal Chem, 2016. **2016**: p. 7436849.
4. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down proteomics: facts and perspectives.* Biochem Biophys Res Commun, 2014. **445**(4): p. 683-93.
5. Siuti, N., et al., *Gene-specific characterization of human histone H2B by electron capture dissociation.* J Proteome Res, 2006. **5**(2): p. 233-9.
6. Zhang, J., et al., *Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure.* J Proteome Res, 2011. **10**(9): p. 4054-65.
7. Barrera, N.P., et al., *Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions.* Nat Methods, 2009. **6**(8): p. 585-7.
8. Mohr, J., et al., *High-efficiency nano- and micro-HPLC--high-resolution Orbitrap-MS platform for top-down proteomics.* Proteomics, 2010. **10**(20): p. 3598-609.
9. Eeltink, S., et al., *1 mm ID poly(styrene-co-divinylbenzene) monolithic columns for high-peak capacity one- and two-dimensional liquid chromatographic separations of intact proteins.* J Sep Sci, 2009. **32**(15-16): p. 2504-9.
10. Vellaichamy, A., 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**(4): p. 1234-44.
11. Parks, B.A., et al., *Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers.* Anal Chem, 2007. **79**(21): p. 7984-91.
12. Hamler, R.L., et al., *A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification.* Proteomics, 2004. **4**(3): p. 562-77.
13. Alpert, A.J., *Hydrophobic interaction chromatography of peptides as an alternative to reversed-phase chromatography.* J Chromatogr, 1988. **444**: p. 269-74.
14. Alpert, A.J. and P.C. Andrews, *Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica.* J Chromatogr, 1988. **443**: p. 85-96.
15. Roth, M.J., et al., *Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry.* Anal Chem, 2011. **83**(24): p. 9586-92.
16. Zhang, J., et al., *Proteoform analysis of lipocalin-type prostaglandin D-synthase from human cerebrospinal fluid by isoelectric focusing and superficially porous liquid chromatography with Fourier transform mass spectrometry.* Proteomics, 2014. **14**(10): p. 1223-31.

17. Skinner, O.S., et al., *Native GELFrEE: a new separation technique for biomolecular assemblies.* Anal Chem, 2015. **87**(5): p. 3032-8.
18. Zhang, J., et al., *Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography.* Anal Chem, 2013. **85**(21): p. 10377-84.
19. Riley, N.M., M.S. Westphall, and J.J. Coon, *Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins.* Anal Chem, 2015. **87**(14): p. 7109-16.
20. Patrie, S.M., et al., *Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa.* J Am Soc Mass Spectrom, 2004. **15**(7): p. 1099-108.
21. Cleland, T.P., et al., *High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer.* J Proteome Res, 2017. **16**(5): p. 2072-2079.
22. Fort, K.L., et al., *Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics.* Anal Chem, 2016. **88**(4): p. 2303-10.
23. Mann, M., R.C. Hendrickson, and A. Pandey, *Analysis of proteins and proteomes by mass spectrometry.* Annu Rev Biochem, 2001. **70**: p. 437-73.
24. Zhai, H., et al., *Consecutive ion activation for top down mass spectrometry: improved protein sequencing by nozzle-skimmer dissociation.* Anal Chem, 2005. **77**(18): p. 5777-84.
25. Yamada, N., E. Suzuki, and K. Hirayama, *Effective novel dissociation methods for intact protein: heat-assisted nozzle-skimmer collisionally induced dissociation and infrared multiphoton dissociation using a Fourier transform ion cyclotron resonance mass spectrometer equipped with a micrometal electrospray ionization emitter.* Anal Biochem, 2006. **348**(1): p. 139-47.
26. Loo, J.A., et al., *High-resolution tandem mass spectrometry of large biomolecules.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 286-9.
27. Coon, J.J., *Collisions or electrons? Protein sequence analysis in the 21st century.* Anal Chem, 2009. **81**(9): p. 3208-15.
28. Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry.* J Mass Spectrom, 2004. **39**(10): p. 1091-112.
29. McLafferty, F.W., et al., *Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance.* J Am Soc Mass Spectrom, 2001. **12**(3): p. 245-9.
30. Macek, B., et al., *Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer.* Mol Cell Proteomics, 2006. **5**(5): p. 949-58.
31. Bondarenko, P.V., et al., *Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer.* J Am Soc Mass Spectrom, 2009. **20**(8): p. 1415-24.

32. Zhang, J., H. Liu, and V. Katta, *Structural characterization of intact antibodies by high-resolution LTQ Orbitrap mass spectrometry.* J Mass Spectrom, 2010. **45**(1): p. 112-20.
33. Fornelli, L., et al., *Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS.* Mol Cell Proteomics, 2012. **11**(12): p. 1758-67.
34. LeDuc, R.D., et al., *ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W340-5.
35. Taylor, G.K., et al., *Web and database software for identification of intact proteins using "top down" mass spectrometry.* Anal Chem, 2003. **75**(16): p. 4081-6.
36. Meng, F., et al., *Informatics and multiplexing of intact protein identification in bacteria and the archaea.* Nat Biotechnol, 2001. **19**(10): p. 952-7.
37. Zamdborg, L., et al., *ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W701-6.
38. Horn, D.M., R.A. Zubarev, and F.W. McLafferty, *Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules.* J Am Soc Mass Spectrom, 2000. **11**(4): p. 320-32.
39. Zabrouskov, V., et al., *New and automated MSn approaches for top-down identification of modified proteins.* J Am Soc Mass Spectrom, 2005. **16**(12): p. 2027-38.
40. Vassall, K.A., V.V. Bamm, and G. Harauz, *MyelStones: the executive roles of myelin basic protein in myelin assembly and destabilization in multiple sclerosis.* Biochem J, 2015. **472**(1): p. 17-32.
41. Baron, W. and D. Hoekstra, *On the biogenesis of myelin membranes: sorting, trafficking and cell polarity.* FEBS Lett, 2010. **584**(9): p. 1760-70.
42. Musse, A.A. and G. Harauz, *Molecular "negativity" may underlie multiple sclerosis: role of the myelin basic protein family in the pathogenesis of MS.* Int Rev Neurobiol, 2007. **79**: p. 149-72.
43. Boggs, J.M., *Myelin basic protein: a multifunctional protein.* Cell Mol Life Sci, 2006. **63**(17): p. 1945-61.
44. Harauz, G., et al., *Myelin basic protein-diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis.* Micron, 2004. **35**(7): p. 503-42.
45. Constantinescu, C.S., et al., *Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS).* Br J Pharmacol, 2011. **164**(4): p. 1079-106.
46. Piaton, G., et al., *Remyelination in multiple sclerosis.* Prog Brain Res, 2009. **175**: p. 453-64.
47. Poser, C.M. and V.V. Brinar, *Diagnostic criteria for multiple sclerosis.* Clin Neurol Neurosurg, 2001. **103**(1): p. 1-11.
48. Frohman, E.M., et al., *Characterizing the mechanisms of progression in multiple sclerosis: evidence and new hypotheses for future directions.* Arch Neurol, 2005. **62**(9): p. 1345-56.

49. Kobelt, G., et al., *Costs and quality of life in multiple sclerosis: a cross-sectional study in the United States.* Neurology, 2006. **66**(11): p. 1696-702.
50. Pelletier, D. and D.A. Hafler, *Fingolimod for multiple sclerosis.* N Engl J Med, 2012. **366**(4): p. 339-47.
51. Ransohoff, R.M., *Natalizumab for multiple sclerosis.* N Engl J Med, 2007. **356**(25): p. 2622-9.
52. Bielekova, B., et al., *Encephalitogenic potential of the myelin basic protein peptide (amino acids 83-99) in multiple sclerosis: results of a phase II clinical trial with an altered peptide ligand.* Nat Med, 2000. **6**(10): p. 1167-75.
53. Hafler, D.A., *Multiple sclerosis.* J Clin Invest, 2004. **113**(6): p. 788-94.
54. Miller, D.H., et al., *Differential diagnosis of suspected multiple sclerosis: a consensus approach.* Mult Scler, 2008. **14**(9): p. 1157-74.
55. Noseworthy, J.H., et al., *Multiple sclerosis.* N Engl J Med, 2000. **343**(13): p. 938-52.
56. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis.* Ann Neurol, 2001. **50**(1): p. 121-7.
57. Hawkes, C.H. and G. Giovannoni, *The McDonald Criteria for Multiple Sclerosis: time for clarification.* Mult Scler, 2010. **16**(5): p. 566-75.
58. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria.* Ann Neurol, 2011. **69**(2): p. 292-302.
59. Lassmann, H., W. Bruck, and C. Lucchinetti, *Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy.* Trends Mol Med, 2001. **7**(3): p. 115-21.
60. Polman, C.H., J.S. Wolinsky, and S.C. Reingold, *Multiple sclerosis diagnostic criteria: three years later.* Mult Scler, 2005. **11**(1): p. 5-12.
61. Lim, S.Y. and C.S. Constantinescu, *Current and future disease-modifying therapies in multiple sclerosis.* Int J Clin Pract, 2010. **64**(5): p. 637-50.
62. Berger, T., et al., *Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event.* N Engl J Med, 2003. **349**(2): p. 139-45.
63. Barnett, M.H. and J.W. Prineas, *Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion.* Ann Neurol, 2004. **55**(4): p. 458-68.
64. Benson, J.M., et al., *Oral administration of myelin basic protein is superior to myelin in suppressing established relapsing experimental autoimmune encephalomyelitis.* J Immunol, 1999. **162**(10): p. 6247-54.
65. Frohman, E.M., M.K. Racke, and C.S. Raine, *Multiple sclerosis--the plaque and its pathogenesis.* N Engl J Med, 2006. **354**(9): p. 942-55.
66. Steinman, L. and S.S. Zamvil, *How to successfully apply animal studies in experimental allergic encephalomyelitis to research on multiple sclerosis.* Ann Neurol, 2006. **60**(1): p. 12-21.
67. Baxter, A.G., *The origin and application of experimental autoimmune encephalomyelitis.* Nat Rev Immunol, 2007. **7**(11): p. 904-12.
68. Furlan, R., C. Cuomo, and G. Martino, *Animal models of multiple sclerosis.* Methods Mol Biol, 2009. **549**: p. 157-73.

69. Musse, A.A., J.M. Boggs, and G. Harauz, *Deimination of membrane-bound myelin basic protein in multiple sclerosis exposes an immunodominant epitope.* Proc Natl Acad Sci U S A, 2006. **103**(12): p. 4422-7.
70. Lafaille, J.J., et al., *Myelin basic protein-specific T helper 2 (Th2) cells cause experimental autoimmune encephalomyelitis in immunodeficient hosts rather than protect them from the disease.* J Exp Med, 1997. **186**(2): p. 307-12.
71. Nave, K.A. and H.B. Werner, *Myelination of the nervous system: mechanisms and functions.* Annu Rev Cell Dev Biol, 2014. **30**: p. 503-33.
72. Han, M.H., et al., *Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets.* Nature, 2008. **451**(7182): p. 1076-81.
73. Lock, C., et al., *Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis.* Nat Med, 2002. **8**(5): p. 500-8.
74. Kornek, B., et al., *Multiple sclerosis and chronic autoimmune encephalomyelitis: a comparative quantitative study of axonal injury in active, inactive, and remyelinated lesions.* Am J Pathol, 2000. **157**(1): p. 267-76.
75. Kornek, B. and H. Lassmann, *Neuropathology of multiple sclerosis-new concepts.* Brain Res Bull, 2003. **61**(3): p. 321-6.
76. Paz Soldan, M.M. and M. Rodriguez, *Heterogeneity of pathogenesis in multiple sclerosis: implications for promotion of remyelination.* J Infect Dis, 2002. **186 Suppl 2**: p. S248-53.
77. Biber, A., et al., *Myelin basic protein in cerebrospinal fluid of patients with multiple sclerosis and other neurological diseases.* J Neurol, 1981. **225**(4): p. 231-6.
78. Kim, J.K., et al., *Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis.* Mol Cell Proteomics, 2003. **2**(7): p. 453-62.
79. Artemiadis, A.K. and M.C. Anagnostouli, *Apoptosis of oligodendrocytes and post-translational modifications of myelin basic protein in multiple sclerosis: possible role for the early stages of multiple sclerosis.* Eur Neurol, 2010. **63**(2): p. 65-72.
80. Mastronardi, F.G., et al., *Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated.* J Neurosci Res, 2007. **85**(9): p. 2006-16.
81. Moscarello, M.A., et al., *Myelin in multiple sclerosis is developmentally immature.* J Clin Invest, 1994. **94**(1): p. 146-54.
82. Nave, K.A., *Neurological mouse mutants and the genes of myelin.* J Neurosci Res, 1994. **38**(6): p. 607-12.
83. Moscarello, M.A., F.G. Mastronardi, and D.D. Wood, *The role of citrullinated proteins suggests a novel mechanism in the pathogenesis of multiple sclerosis.* Neurochem Res, 2007. **32**(2): p. 251-6.
84. Pritzker, L.B., et al., *Deimination of myelin basic protein. 2. Effect of methylation of MBP on its deimination by peptidylarginine deiminase.* Biochemistry, 2000. **39**(18): p. 5382-8.
85. Mastronardi, F.G., et al., *Increased citrullination of histone H3 in multiple sclerosis brain and animal models of demyelination: a role for tumor necrosis factor-induced peptidylarginine deiminase 4 translocation.* J Neurosci, 2006. **26**(44): p. 11387-96.

CHAPTER TWO

SUPERFICIALLY POROUS REVERSED PHASE LIQUID CHROMATOGRAPHY

This chapter was published in Analytical Chemistry: Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry. Anal Chem. 2011 Dec 15;83(24):9586-92. doi: 10.1021/ac202339x. Epub 2011 Nov 15.

Abstract

The compatibility of superficially porous (SP) resin for label-free intact protein analysis with on-line capillary LC/MS is demonstrated to give improved chromatographic resolution, sensitivity, and reproducibility. The robustness of the platform was measured against several samples of varying complexity and sample loading amount. The results indicate that capillary SP columns provide high loading capacities and that ~6 second chromatographic peak widths are typical for standard proteins in simple mixtures and proteins isolated from cell and tissue lysates. Subfemtomole detection limits for standard proteins were consistently observed, with the lowest levels at 12 amol for ubiquitin. The analysis of total heart homogenates shows that capillary SP columns provide theoretical peak capacity of 106 protein forms with 30 minute total analysis time, and enabled detection of proteins from complex mixtures with a single high-resolution scan. The SPLC/MS platform also detected 343 protein forms from two HeLa acid extract replicate analyses that consumed 5×10^4 cells and 30 minutes analysis time, each. Comparison of all the species observed in each HeLa replicate showed 90% overlap (309 forms) with a Pearson correlation coefficient of 89.9%

for the common forms observed in the replicates. Efficient acid extract of 1 x 10^4 HeLa cells allowed reproducible detection of common modification states and members from all five of the histone families, and demonstrated that capillary SPLC/MS supports reproducible label-free profiling of histones in <15 minutes total analysis time. The data presented demonstrate that a capillary LC/MS platform utilizing superficially porous stationary phase and a LTQ-Orbitrap FT-MS is fast, sensitive, and reproducible for intact protein profiling from small tissue and cell amounts.

Introduction

The analysis of intact proteins with top-down MS has emerged in the last decade as a powerful method for monitoring the variety of biological events that occur at the gene, transcript, and protein levels [1]. A recent review describes increased implementation of top-down MS for both targeted and discovery modes of analysis [2]. While robust data acquisition and informatics tools have been developed for the generation and characterization of top-down datasets, identification of hundreds of endogenous proteins from complex mixtures still requires days of dedicated instrument time using existing tools [3, 4]. This is largely because multidimensional separations are used to improve dynamic range, but at a cost of generating tens to hundreds of fractions that burden the experimental duty cycle for traditional LC/MS platforms [4, 5]. With the growing use of top-down MS as a technique, improvements in separations, MS hardware, and automation are required to provide reliable information on an efficient time scale.

A summary of the chromatographic and mass spectrometric methods for top-down proteomics was recently published by Capriotti, *et al.* [6]. Top-down MS is often performed with reversed-phase liquid chromatography (RPLC) coupled directly to high-performance MS hardware (*i.e.*, LC/MS). Even with pre-fractionation steps separation of intact proteins with RPLC typically necessitates long-gradient durations (>100 minutes) because the broad physicochemical diversity of proteins, their post-translational modifications, and the potential for isomeric forms with differing exposed residues reduces the column peak capacity. Various reversed-phase resins with varying column sizes and stationary phases have been utilized for online LC/MS of intact proteins. Reports on the use of monolithic [7, 8], polymeric [4], porous silica [5], and nonporous silica [9] as RPLC phases in top-down LC/MS demonstrate varying figures of merit for chromatographic resolution, speed, and sensitivity. The nature of the various stationary phases used for intact protein LC/MS directly impacts the performance of key metrics as summarized in Table S-2 (Appendix I) [10-12]. Of these, superficially porous (SP) RPLC resin, in which a nonporous silica core is coated with a thin layer of porous silica, has been shown to have mass transfer efficiencies similar to those of nonporous silica while maintaining high loading capacities at back pressures consistent with standard HPLC pumping systems (*i.e.*, no need for ultrahigh pressure liquid chromatography (UPLC) equipment) [13]. These characteristics enable analysis at high linear velocities for fast separations with minimal increases in plate height [14]. and has been shown to allow fast separation of simple mixtures of intact proteins using conventional HPLC equipment [15].

Among the various top-down MS platforms published to date, high-performance Fourier transform ion cyclotron resonance (FT-ICR) MS stands as the most often used mass analyzer due to the high-resolution and mass accuracy provided at high magnetic field strengths [6]. Many have recognized that the high sampling rates and sensitivities of lower resolution instruments facilitate high-throughput tandem MS experiments on intact proteins [16]. For example, the improving performance of commercial quadrupole/time-of-flight (Q-TOF) MS systems (40000 resolution at 400 *m/z*) and Orbitrap MS (100000 resolution at 400 *m/z*) [17] makes possible top-down MS on a faster time scale than is possible with FT-ICR MS. Given the characteristics of good-mass resolution, sensitivity, scan rate, and its wide-spread application for peptide-based applications, the Orbitrap mass spectrometer is a good choice for development of a robust on-line LC/MS platform for intact protein analysis.

We present here a new LC/MS platform that combines the high performance Orbitrap MS with capillary SP RPLC for the analysis of intact proteins (denoted SPLC/MS). The chromatographic resolution and speed provided by capillary SP RPLC challenges the scanning speed of the Orbitrap, and provides improved limits of detection (LOD) for intact proteins compared to conventional capillary HPLC stationary phases. Importantly, the robustness, reproducibility and sensitivity of the SPLC/MS platform is evaluated with complex protein mixtures derived from heterogeneous tissue homogenates and cell lysates from $<1x10^5$ cell counts.

Methods

Reagents and Materials

All reagents and protein standards were purchased from Sigma (St. Louis, MO). Solvents and acids used for LC/MS were Optima Grade (Thermo Scientific, Waltham, MA). Mouse heart homogenate and HeLa acid extracts were prepared as described in supporting information.

Capillary SP RPLC Columns

Capillary RPLC columns were packed with C18 Poroshell-300 (Agilent Technologies, Santa Clara, CA) particles, 5 μm diameter, 300 Å pore size. Resin was packed (either in-house or by New Objective, Inc., Woburn, MA) to a bed length of 15 cm into Picofrit columns (New Objective, Inc.) 75 μm I.D. x 360 μm O.D. with a terminal 15 μm I.D. μESI spray tip. Picofrit columns contain integral frits and microspray tips to minimize extra-column broadening.

Online LC/MS

Standard protein or desalted complex endogenous mixtures were injected into a 1 μL sample loop on an 1100 nano-LC (Agilent, Santa Clara, CA) maintained at a 1 μL/min flow rate. The column was maintained at 60 ^{o}C by a capillary column heater (Analytical Sales & Services, Pompton Plains, NJ). Mobile phases for RPLC were as follows: solvent A 0.025% TFA, 0.3% formic acid, 20% acetonitrile in water; solvent B 0.025% TFA, 0.3% formic acid, 20% isopropanol in acetonitrile. Standard proteins were analyzed with a 3 minute linear elution

gradient (0-40% B in 3 minutes, 40-90% B in 4 minutes). For complex mixtures a samples were desalted at 0% B for 5 minutes and longer elution gradient (*e.g.,* 5 or 20 minutes) were used. The Picofrit/ESI emitter was interfaced to the LTQ Orbitrap XL or LTQ Orbitrap Velos (Thermo Fisher, Waltham, MA) capillary inlet with a custom μESI source. Mass spectrometers were tuned for the $[M+6H]^{6+}$ of ubiquitin by direct infusion with the custom μESI source. Mass spectrometer settings and data analysis methods are described in supporting information.

Results and Discussion

Detection Limits and Analytical Figures of Merit for the Analysis of Protein Standards.

Detection limits for standard proteins. Reducing column dimensions in LC/MS improves detection limits for proteins and peptides [8]. Previous top-down studies that integrated capillary RPLC and high-performance MS reported protein LOD from a few to a few hundred femtomoles [4, 7]. These values are above those of other common non-immunoassay protein detection methods; for example, polyacrylamide gel electrophoresis (PAGE) with silver staining (~30-200 fmol) [18], multiple reaction monitoring (MRM) (10-50 amol) [19], or bottom-up protocols such as peptide mass fingerprinting (0.5-5 fmol) [20, 21]. The sensitivities of the previous online LC/MS top-down platforms were limited by relatively poor RPLC peak resolution, with reported peak widths at half max (w_h) of 17-60 seconds. These peak widths far exceed the temporal resolution of individual MS scan events (typically 0.1-1.5 seconds) with detrimental effects on experimental LOD and peak capacity.

Kirkland previously showed that SP "Poroshell" resins, consisting of a solid core and <1 μm porous outer shell, allow fast protein separations with w_h of a few seconds [15, 22]. This work seeks to leverage the advantages of SP resins for on-line capillary LC and high-resolution MS for analysis of intact proteins. Performance metrics were initially evaluated by separation of equimolar mixtures of bovine ubiquitin, chicken lysozyme, bovine serum albumin (BSA), horse heart myoglobin, and bovine carbonic anhydrase with a three minute LC gradient and a LTQ Orbitrap XL. The expanded total ion current (TIC) chromatogram for a 20 fmol injection (Figure 1a), analyzed at a resolution of 15000 (at 400 *m/z*), demonstrates fast, high resolution separations with w_h from 2.5-5 seconds. The good chromatographic resolution led to strong mass spectral signal response and low LOD for each protein. For example, mass spectra at a resolution of 60000 (at 400 *m/z*) of each protein in a 1 fmol injection showed S/N ratios >5 (Figure 1b) with calculated LOD from 34 to 536 amol (Table S-1, Appendix I). These data indicated intact protein LOD with SPLC/MS are similar to those achieved by PAGE with silver staining [18] and MRM analyses [19] for these standard proteins.

The results for a 1 fmol injection were corroborated on a separate LTQ-Orbitrap Velos, where calculated LOD <154 amol were obtained for most of the proteins (Table 1). The LTQ-Orbitrap Velos provided 3-12× lower LOD for ubiquitin, lysozyme, and myoglobin versus the LTQ-Orbitrap XL. For example, Figure S-1 (Appendix I) shows spectra with comparable S/N obtained at 100 amol on the LTQ-Orbitrap XL and 20 amol on the LTQ-Orbitrap Velos. These results are in agreement with previous efforts that show a 3-5×

improvement in ion transfer efficiency for the LTQ-Orbitrap Velos compared to the LTQ-Orbitrap XL instruments [23]. Interestingly, the LOD for carbonic anhydrase showed minor

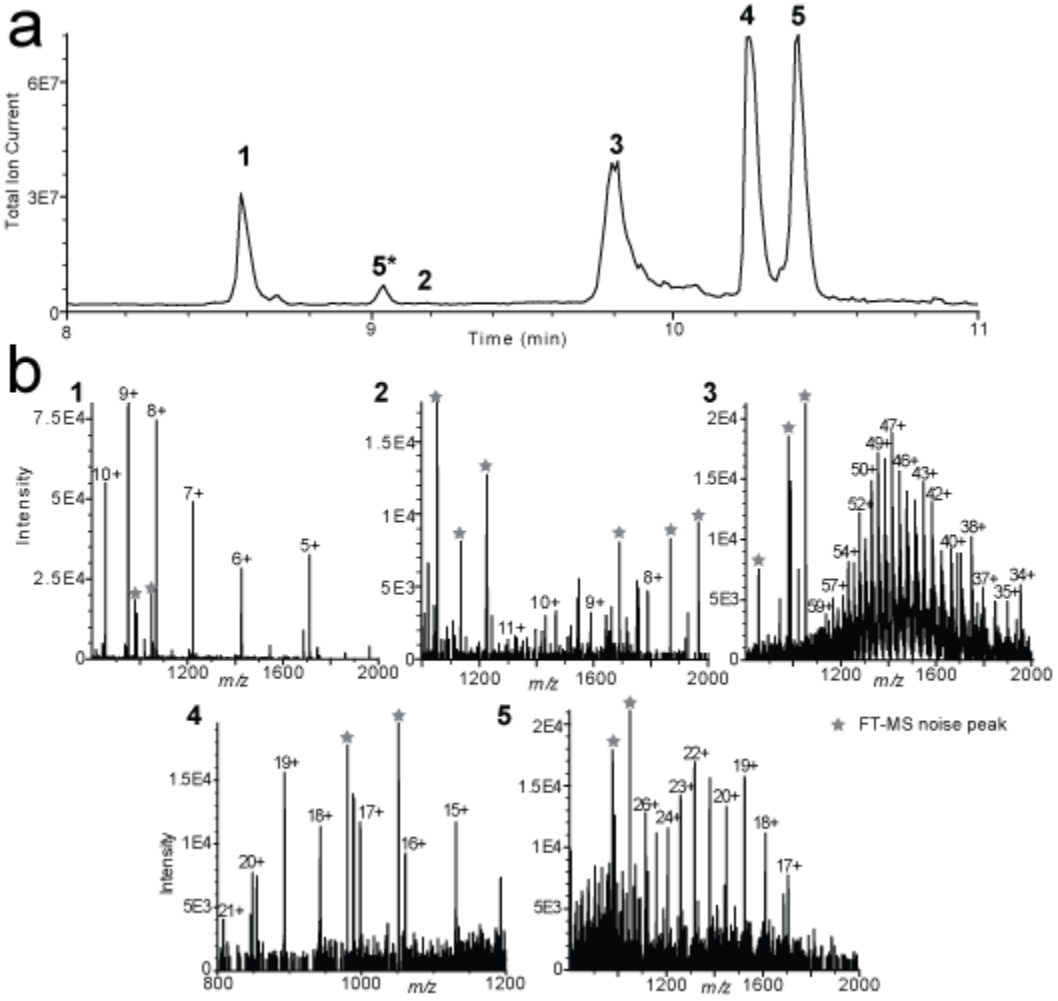


Figure 1. Sensitive SPLC/MS detection of standard protein mixture. a) Total ion current (TIC) chromatogram demonstrates separation of the five components with a 3 min elution gradient for 20 fmol loaded. b) Representative mass spectra demonstrate S/N ranges from 5-87 for the individual components observed in the equimolar mixture (1 fmol/constituent on column), with S/N, LOD, and LLOQ information presented in Table 1. Components: 1) bovine ubiquitin, 2) chicken lysozyme, 3) bovine serum albumin, 4) equine myoglobin, 5) bovine carbonic anhydrase, 5*) 15.6 kDa contamination in carbonic anhydrase, likely superoxide dismutase [24].

change and BSA was not detected with the LTQ-Orbitrap Velos. Previous work with "ion funnel" technologies showed *m/z* biases [23] and similar reduced transmission of higher mass species for the stacked ring radio frequency (RF) ion guide at elevated pressures [25]. Many commercial mass spectrometers, like the LTQ-Orbitrap Velos, improve ion capture efficiency via ion funnel technology at the capillary interface; however our current findings and these previous reports indicate that for intact protein mixtures care must be taken when tuning the instrument to avoid *m/z* and mass bias.

Label-free Quantitation. The dynamic range of the SPLC/MS platform was assessed by a dilution series from 50 amol to 200 fmol on three unique capillary columns. Calibration curves plotting extracted ion chromatogram (EIC) peak area for the $[M+6H]^{6+}$ to $[M+10H]^{10+}$ ubiquitin ions (Figure 2a), illustrate quantitative signal increases across a linear dynamic range of 4000 (50 amol to 200 fmol). These data show subfemtomole LOD are not limited to a single column, but are generally achievable with the capillary SP-RPLC combined with the LTQ-Orbitrap XL. Additionally, the strong linear correlation indicates that this platform may be capable of rapid label-free relative quantitation across samples [26].

Column loading capacity. Excellent peak capacity and rapid analysis of intact proteins by capillary LC/MS has previously been demonstrated using HPLC with monolithic polymer columns [7, 8, 27] and UPLC with non-porous silica and polymer columns [28]. Previous work demonstrated column loading capacities of 50 pg for a standard protein on 200 μm x 5 cm monolithic polymer columns [12]. For the SPLC/MS platform, a plot of peak width as a

function of amount of ubiquitin loaded demonstrates ~5 second w_h from 0.8 pg to 4.3 ng (Figure 2b), indicating loading capacities of ~5 ng for ubiquitin can be achieved on 75 μm I.D. columns without significant peak broadening. The increased protein loading capacity and dynamic range for SPLC/MS compared to reported metrics for monolithic columns is in agreement with results for small molecule HPLC analysis [10].

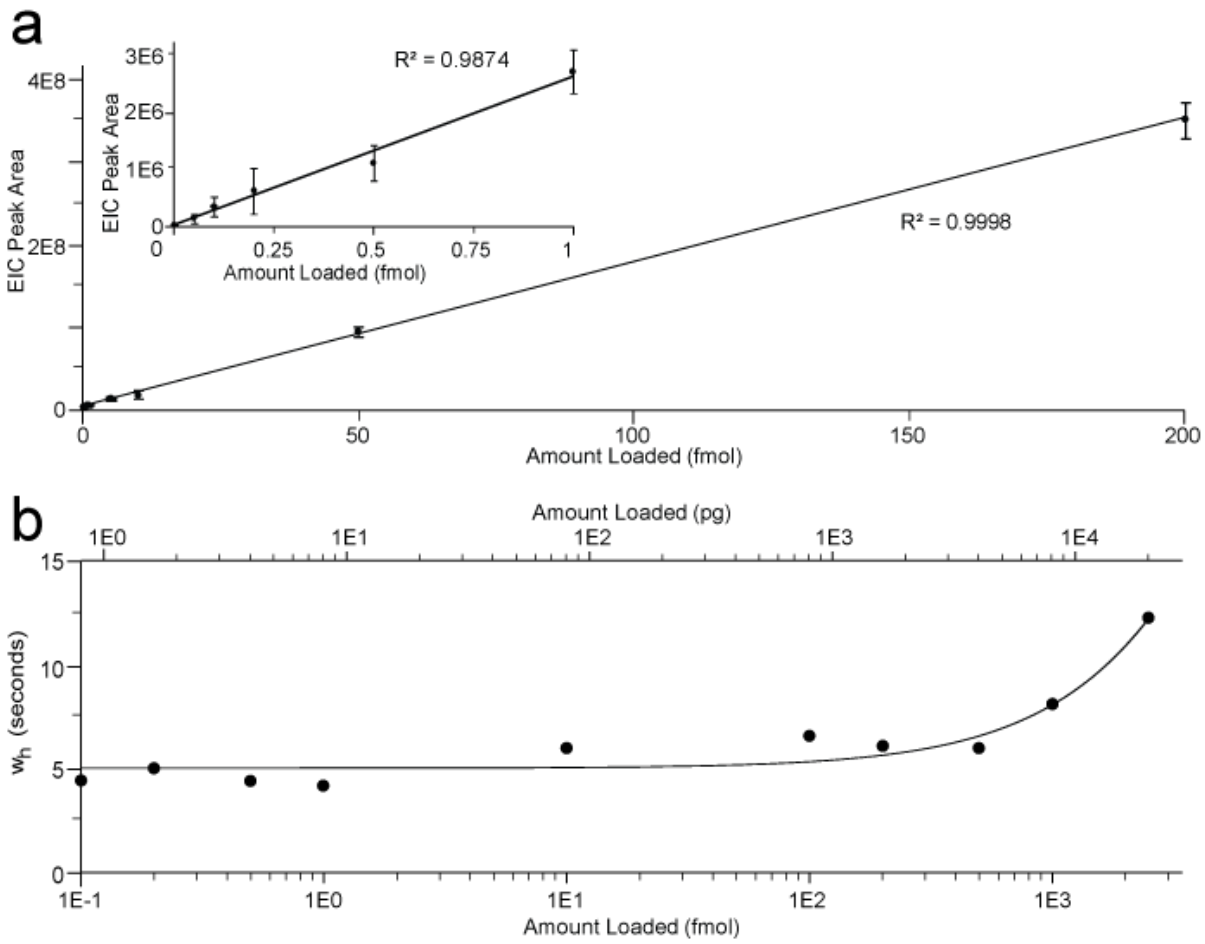


Figure 2. Column-to-column ubiquitin MS signal variation and peak width at various injection loads. a) Averaged calibration curve from three unique columns shows linear detection of ubiquitin from 50 amol to 200 fmol with expanded view of 50-1000 amol (inset). b) Peak width as a function of ubiquitin loaded demonstrating consistent peak width up to 5 ng.

Reproducibility at LLOQ. Analytical performance metrics at the lower limit of quantitation (LLOQ) were established with 25 replicate injections of 200 amol ubiquitin (1.75× LLOQ). An expanded view of the EIC for ubiquitin ions $[M+6H]^{6+}$ to $[M+10H]^{10+}$ demonstrate good agreement in peak height and width over the 25 replicates (Figure 3a). In these experiments carbon-12/carbon-13 isotope statistics were readily obtained via Xtract and matched to theoretical values with 3 ppm mass accuracy (Figure 3b). Estimates of error for these

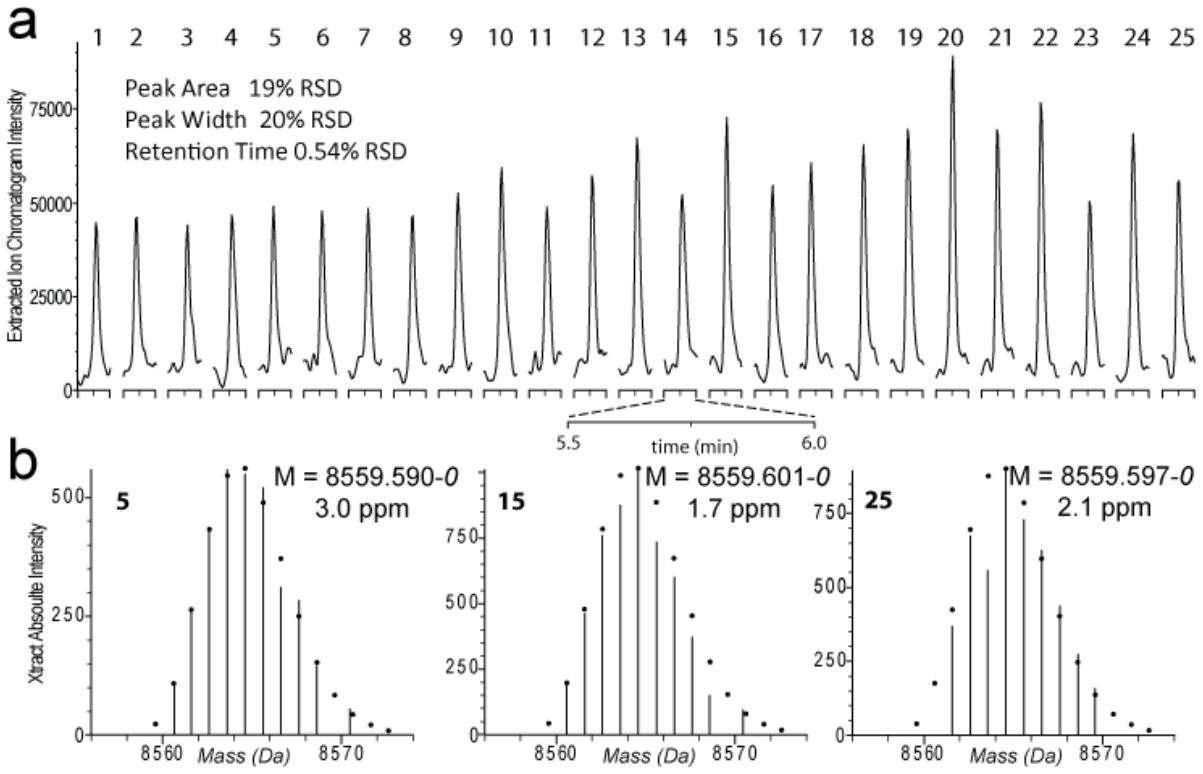


Figure 3. Repeated measurement of 200 amol ubiquitin. a) Extracted ion chromatograms (EIC) for the time interval 5.5-6.0 minutes from each of 25 consecutive replicates demonstrate reproducible ubiquitin detection. b) Mass spectra for the fifth, fifteenth, and twenty-fifth replicates compares the observed decharged $^{12}C/^{13}C$ isotopic distribution, determined by Xtract for the average of 6 spectra, with the theoretical isotopic distribution (shown by dots). In each case the observed isotopic distributions yielded monoisotopic masses within 3 ppm of predicted (see inset).

different metrics show that peak width and area varied by 19% RSD and 20% RSD, respectively (Table S-3, Appendix I), while variation in the deisotoped, monoisotopic Xtract peak height was 40% RSD. Presumably, some of this error derives from the sample preparation at low amol levels. The increased variation associated with Xtract peak intensity relative to the peak area and width presumably comes from incomplete calculated isotopic envelopes generated during *z* and mass determination with Xtract (*e.g.*, Figure 3b, replicate 25). Also, we estimate that ~20% of the observed variance comes from artificial peak broadening associated with limited scanning speed for high-resolution scan events (~6 scans events per ubiquitin peak) across the chromatographic peak [7]. Interfacing capillary SPLC to faster MS systems such as Q-TOF or using enhanced FT-MS modes [29-34] will provide improved error tolerance due to the greater number of data points generated across the chromatographic peak at higher duty cycle.

These results demonstrate that the peak area of the EIC provided the most reproducible means by which to quantitate for these replicates, although the Xtract peak height demonstrated significant correlation as well. Given the difficulties with automating EIC peak area calculations and the speed of implementing Xtract calculations, Xtract peak height was chosen as the method for quantitation in subsequent complex mixture analyses. These data suggest that future efforts to fit LC/MS datasets with statistical models for label-free quantitative proteomics using differential mass spectrometry (DMS) [26, 35, 36] would benefit from the speed and reproducibility of the SPLC/MS platform, providing reproducible intact protein DMS information in less time.

LC/MS Profiling of Complex Mixtures.

Analysis of mouse total heart homogenate. Label-free differential intact protein profiling has been discussed previously [26] and is commonly used for comparison of protein isoforms from simple mixtures (*e.g.*, cholesterol particles). Established metrics are limited on single dimension separations for label-free analysis of complex mixtures that have thousands of proteins present over broad concentration ranges (*e.g.*, tissue homogenates). Analysis of such a mixture is often performed using 10-1000's µg of material and multidimensional separations to increase the total peak capacity and improve dynamic range [1, 3-5, 37]. Single dimension theoretical peak capacity (n_c) for HPLC has previously been described [38].

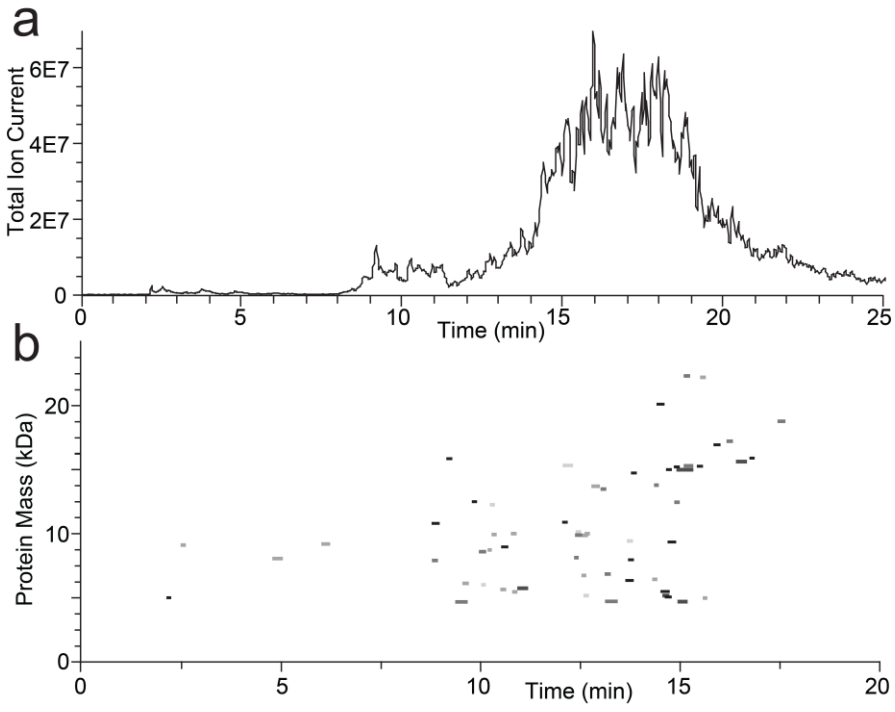


Figure 4

Figure 4. SPLC/MS analysis of crude mouse whole heart homogenate with a 20 minute elution gradient. a) The TIC shows a complex chromatogram with 59 proteineacious species determined by Xtract. b) Virtual 2D gel showing SP-RPLC retention time vs. protein molecular weight for 59 unique protein species.

Traditional HPLC detectors necessitate well resolved species for n_c estimates, however, multiplexed detection with MS provides an opportunity to estimate a practical n_c value for complex mixtures when protein species are not chromatographically resolved. The n_c of the SPLC/MS platform was evaluated by separation of ~500 ng total protein from mouse heart homogenate with a 20 minute gradient. Previous studies show that heart homogenates contain 1000's of proteins expressed over a large concentration range [39]. With SPLC/MS, a complex total ion current chromatogram was observed from the heart homogenate (Figure 4a). Analysis of individual mass spectral scan events with Xtract determined 59 unique, manually validated, proteinaceous species with masses from 4.6-22.3 kDa (Figure 4b). For the 59 components detected with Xtract the most intense component, representing the broadest peak width of any component observed, had a w_h of 21.6 seconds, however, the mean and median w_h for the 59 species were 6.8 seconds and 5.4 seconds, respectively, which generally agree with results from standard proteins (*v.s.*).

For the 20 minute RPLC gradient used to analyze the mouse total heart homogenate, a mean n_c of 106 intact protein forms was determined for the SPLC/MS experiment. Greater peak capacities have been observed using other RPLC platforms on time scales similar to that shown here, although these values were generated from standard proteins rather than a complex mixture of endogenous proteins with observed protein numbers that approach the

theoretical peak capacity [8, 27, 40]. The data suggest that combining the speed and peak capacity of SPLC/Orbitrap platform with an orthogonal separation method, such as isoelectric focusing or GELFrEE [41, 42], on 0.5-5 mg of starting material will dramatically increase the number of intact proteins observed [43].

Reproducible LC/MS profiling of HeLa acid extract at low cell count. The SPLC/MS platform was used to profile acid soluble fractions of two unique HeLa S3 cell pellets, corresponding to ~5 x 10^4 cells per analysis. Samples were analyzed using a 20 minute linear gradient with a total analysis time of 30 minutes. From individual LC/MS scans, a total of

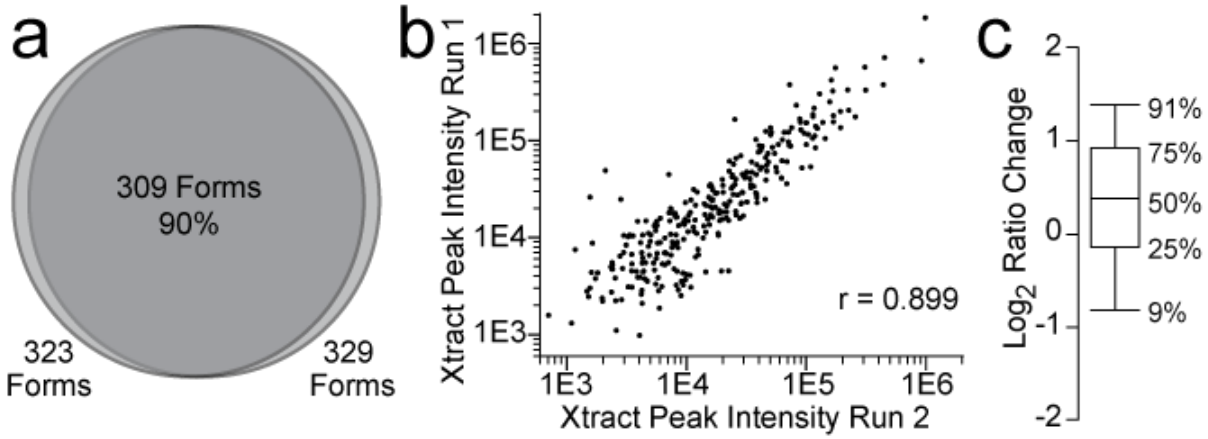


Figure 5

Figure 5. Comparison of two HeLa acid extract replicates (~5 x 10^4 cells) with SPLC/MS; 323 and 329 protein species were observed for the separate SPLC/MS runs with a 20 minute elution gradient. a) The Venn diagram illustrates 90% overlap within a 1 Da Δmass window between the replicates. b) A scatter plot compares monoisotopic Xtract peak intensity for the common species. c) A Log_2 y/x box-and-whisker plot illustrates the typical spread of similar HeLa acid extract replicates. Boxes represent 1st quartile, median, and 3rd quartile; whiskers represent 5th and 95th percentiles, respectively.

343 validated, non-redundant protein forms were detected in individual MS scans from the two separate analyses. For the observed species the deconvoluted monoisotopic peak intensities ranged from 7.0E2 to 1.9E6 and gave a dynamic range estimate of 2700 for the complex mixture analysis. A total of 309 common forms were observed in both runs, yielding >90% overlap among the replicates (Figure 5a). The protein forms not observed in both runs were at ≤1.5% abundance relative to the most intense species, which indicates the most intense species were reproducibly observed among the replicates. A scatter plot of signal intensities for the common protein forms in the replicates shows strong correlation between the replicates (Figure 5b) with a Pearson correlation coefficient of 89.9%. The box plot of the observed ratio between the runs shows 95% of the common forms observed within ~3× intensity change (Figure 5c), illustrating reproducible signal intensities between analyses. Overall, the data indicate that the SPLC/MS platform provides reliable data for intact protein profiling, even from Xtract spectral intensities, and shows that large scale, rapid analysis of complex mixtures with existing mass spectrometric and chromatographic tools is possible with SP resins. The fact that each of these proteins was detected from a single SPLC/MS scan highlights quality isotopic envelopes >3:1 S/N were generated throughout the run. Manual inspection of the datasets revealed 10 protein forms from 30-50 kDa that were not detected by Xtract because of unresolved isotopic envelopes. Although not shown here, improved MS information from larger species (>30 kDa) can be obtained using low resolution Orbitrap scans (7500 at 400 *m/z*) or ion trap MS detection [44]. Also, signal averaging prior to analysis by Xtract provides improved isotopic statistics [45] but is not routinely possible with current Xcaliber software, highlighting the need for enhanced top-

down data processing platforms designed for FT-ICR instrumentation to be extended to the Orbitrap [46].

Histone Profiling: Analysis of histone H2A, H2B and H4 proteins by MS has previously been demonstrated for small sample amounts (5 x 10^4 cells) using subcellular fractionation and long (100 minute) RPLC gradients [47]. For the SPLC/MS runs performed on 5 x 10^4 Hela cells (*v.s.*), the intact mass tags for major classes of histones were easily discerned including H3.1 (Supplementary Figure 2a, Appendix), which previously challenged MS analysis at the same cell count [47]. Notably, these results were generated in the absence of nucleus isolation which highlights the good peak capacity and dynamic range of the SPLC/MS platform even when challenged with complex mixtures that are traditionally simplified by subcellular fractionation.

We further evaluated SPLC/MS speed, sensitivity, and reproducibility with three acid extract technical replicates, with 1 x 10^4 cells and a 5 minute SPLC gradient. The MS results yielded mass tags for PTM isoforms and gene family members for histones H1, H2A, H2B (Figure 6a), and H4 (Figure 6b). These results are similar to those reported previously using targeted analyses [48, 49]. As shown for H2B and H4, the triplicate analysis at 1 x 10^4 cells counts yielded reproducible isotopic distribution spectra with average RSD for the PTM/isoform occupancy levels at 13.8% (Figure S-2c and S-2d, Appendix I). The mass spectrum of histone H3.1 was observed with simplified PTM profiles compared to the extraction with 5x10^4 cells (Figure S-2a vs. S-2b, Appendix I); however, the relatively poor

MS data for H3 may be mitigated in the future by minimizing sample dilution prior to loading. Overall, these data illustrate the high degree of reproducibility from limited sample sizes (<10^5 cells) are possible with the SPLC/MS approach which is on par with those previously reported[47] with ~10× faster analysis times.

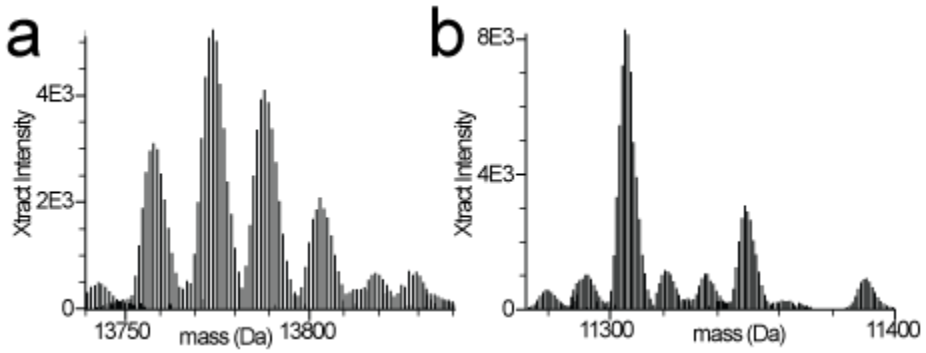


Figure 6. Evaluation of histone classes observed in SPLC/MS of HeLa acid extract. a) Xtract decharged mass spectra for histone H2B protein gene isoforms and b) histone H4 PTM isoforms from 1 x 10^4 HeLa cells.

Conclusions

The SPLC/MS platform provides high chromatographic resolution and speed on traditional HPLC systems with sub-UPLC pressure limits. This report presents the adaptation of SP-RPLC technology to capillary columns with integrated ESI emitters to enable rapid and sensitive analysis of intact proteins from simple and complex mixtures that derive from whole cell or tissue lysates. The data indicate that for various protein standards, reproducible amol LOD and broad linear dynamic range are easily obtained from high resolution MS data. These detection limits are comparable to those commonly observed for peptide mass fingerprinting, traditional silver stained gels, and MRM assays but with less front end sample preparation. For whole cell extracts, strong and reproducible signals for >300 protein forms

were obtained with total analysis times — from cell pellet to MS detection — under two hours. Interassay replicates provided good correlation between signal intensities across large protein concentration ranges and in the presence of highly complex mixtures. Correlation of expression ratios and percent occupancies of histone variants and PTMs between runs shows the method is valuable for protein population analysis. These results demonstrate a platform that combines capillary HPLC with SP stationary phase resin and high-performance MS that is well suited to analysis of intact proteins; and improves throughput and detection limits in intact mass profiling.

References

1. Roth, M.J., et al., *Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry.* Mol Cell Proteomics, 2005. **4**(7): p. 1002-8.
2. Kellie, J.F., et al., *The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput.* Mol Biosyst, 2010. **6**(9): p. 1532-9.
3. Roth, M.J., et al., *"Proteotyping": population proteomics of human leukocytes using top down mass spectrometry.* Anal Chem, 2008. **80**(8): p. 2857-66.
4. Vellaichamy, A., 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**(4): p. 1234-44.
5. Parks, B.A., et al., *Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers.* Anal Chem, 2007. **79**(21): p. 7984-91.
6. Capriotti, A.L., et al., *Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics.* J Chromatogr A, 2011.
7. Mohr, J., et al., *High-efficiency nano- and micro-HPLC--high-resolution Orbitrap-MS platform for top-down proteomics.* Proteomics, 2010. **10**(20): p. 3598-609.
8. Eeltink, S., et al., *1 mm ID poly(styrene-co-divinylbenzene) monolithic columns for high-peak capacity one- and two-dimensional liquid chromatographic separations of intact proteins.* J Sep Sci, 2009. **32**(15-16): p. 2504-9.
9. Hamler, R.L., et al., *A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification.* Proteomics, 2004. **4**(3): p. 562-77.
10. Zheng, J., et al., *Comparison study of porous, fused-core, and monolithic silica-based C18 HPLC columns for Celestoderm-V Ointment analysis.* J Pharm Biomed Anal, 2009. **50**(5): p. 815-22.
11. Guillarme, D., et al., *New trends in fast and high-resolution liquid chromatography: a critical comparison of existing approaches.* Anal Bioanal Chem, 2010. **397**(3): p. 1069-82.
12. Detobel, F., et al., *Parameters affecting the separation of intact proteins in gradient-elution reversed-phase chromatography using poly(styrene-co-divinylbenzene) monolithic capillary columns.* J Chromatogr A, 2010. **1217**(18): p. 3085-90.
13. Cunliffe, J.M. and T.D. Maloney, *Fused-core particle technology as an alternative to sub-2-microm particles to achieve high separation efficiency with low backpressure.* J Sep Sci, 2007. **30**(18): p. 3104-9.
14. Cavazzini, A., et al., *Mass-transfer kinetics in a shell packing material for chromatography.* Anal Chem, 2007. **79**(15): p. 5972-9.
15. Kirkland, J.J., et al., *Superficially porous silica microspheres for fast high-performance liquid chromatography of macromolecules.* J Chromatogr A, 2000. **890**(1): p. 3-13.

16. Siuti, N. and N.L. Kelleher, *Decoding protein modifications using top-down mass spectrometry.* Nat Methods, 2007. **4**(10): p. 817-21.
17. Makarov, A., et al., *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.* Anal Chem, 2006. **78**(7): p. 2113-20.
18. Winkler, C., et al., *Silver- and Coomassie-staining protocols: detection limits and compatibility with ESI MS.* Electrophoresis, 2007. **28**(12): p. 2095-9.
19. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.
20. Bizzarri, M., et al., *A label-free method based on MALDI-TOF mass spectrometry for the absolute quantitation of troponin T in mouse cardiac tissue.* Anal Bioanal Chem, 2008. **391**(5): p. 1969-76.
21. Picariello, G., et al., *Fast screening and quantitative evaluation of internally deleted goat alphas1-casein variants by mass spectrometric detection of the signature peptides.* Rapid Commun Mass Spectrom, 2009. **23**(6): p. 775-87.
22. Kirkland, J.J., *Superficially Porous Silica Microspheres for the Fast High-Performance Liquid-Chromatography of Macromolecules.* Analytical Chemistry, 1992. **64**(11): p. 1239-1245.
23. Olsen, J.V., et al., *A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed.* Mol Cell Proteomics, 2009. **8**(12): p. 2759-69.
24. Armirotti, A., U. Benatti, and G. Damonte, *Top-down proteomics with a quadrupole time-of-flight mass spectrometer and collision-induced dissociation.* Rapid Commun Mass Spectrom, 2009. **23**(5): p. 661-6.
25. Kim, T., et al., *Design and implementation of a new electrodynamic ion funnel.* Anal Chem, 2000. **72**(10): p. 2247-55.
26. Mazur, M.T., et al., *Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry.* Proc Natl Acad Sci U S A, 2010. **107**(17): p. 7728-33.
27. Eeltink, S., et al., *High-resolution separations of protein isoforms with liquid chromatography time-of-flight mass spectrometry using polymer monolithic capillary columns.* J Chromatogr A, 2011. **1218**(32): p. 5504-11.
28. Everley, R.A. and T.R. Croley, *Ultra-performance liquid chromatography/mass spectrometry of intact proteins.* J Chromatogr A, 2008. **1192**(2): p. 239-47.
29. Beu, S.C., et al., *Broadband phase correction of FT-ICR mass spectra via simultaneous excitation and detection.* Anal Chem, 2004. **76**(19): p. 5756-61.
30. Makarov, A., E. Denisov, and O. Lange, *Performance evaluation of a high-field Orbitrap mass analyzer.* J Am Soc Mass Spectrom, 2009. **20**(8): p. 1391-6.
31. Patrie, S.M., et al., *Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa.* J Am Soc Mass Spectrom, 2004. **15**(7): p. 1099-108.
32. Schaub, T.M., et al., *High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla.* Anal Chem, 2008. **80**(11): p. 3985-90.
33. Scigelova, M., et al., *Fourier transform mass spectrometry.* Mol Cell Proteomics, 2011. **10**(7): p. M111 009431.

34. Xian, F., et al., *Automated Broadband Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra.* Anal Chem, 2010. **82**: p. 8807-8812.
35. Meng, F., et al., *Quantitative analysis of complex peptide mixtures using FTMS and differential mass spectrometry.* J Am Soc Mass Spectrom, 2007. **18**(2): p. 226-33.
36. Paweletz, C.P., et al., *Application of an end-to-end biomarker discovery platform to identify target engagement markers in cerebrospinal fluid by high resolution differential mass spectrometry.* J Proteome Res, 2010. **9**(3): p. 1392-401.
37. Patrie, S.M., et al., *Top down mass spectrometry of < 60-kDa proteins from Methanosarcina acetivorans using quadrupole FRMS with automated octopole collisionally activated dissociation.* Mol Cell Proteomics, 2006. **5**(1): p. 14-25.
38. Grushka, E., *Chromatographic peak capacity and the factors influencing it.* Analytical Chemistry, 1970. **42**(11): p. 1142-1147.
39. Chugh, S., C. Suen, and A. Gramolini, *Proteomics and mass spectrometry: what have we learned about the heart?* Curr Cardiol Rev. **6**(2): p. 124-33.
40. Rogeberg, M., et al., *Separation of intact proteins on porous layer open tubular (PLOT) columns.* J Chromatogr A, 2010. **1217**(17): p. 2782-6.
41. Lee, J.E., et al., *A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome.* J Am Soc Mass Spectrom, 2009. **20**(12): p. 2183-91.
42. Tran, J.C. and A.A. Doucette, *Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation.* Anal Chem, 2008. **80**(5): p. 1568-73.
43. Wang, H. and S. Hanash, *Multi-dimensional liquid phase based separations in proteomics.* J Chromatogr B Analyt Technol Biomed Life Sci, 2003. **787**(1): p. 11-8.
44. Durbin, K.R., et al., *Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale.* Proteomics, 2010. **10**(20): p. 3589-97.
45. Compton, P.D., et al., *On the Scalability and Requirements of Whole Protein Mass Spectrometry.* Anal Chem, 2011.
46. Wenger, C.D., et al., *Versatile online-offline engine for automated acquisition of high-resolution tandem mass spectra.* Anal Chem, 2008. **80**(21): p. 8055-63.
47. Li, M., L. Jiang, and N.L. Kelleher, *Global histone profiling by LC-FTMS after inhibition and knockdown of deacetylases in human cells.* J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(30): p. 3885-92.
48. Siuti, N., et al., *Gene-specific characterization of human histone H2B by electron capture dissociation.* J Proteome Res, 2006. **5**(2): p. 233-9.
49. Thomas, C.E., N.L. Kelleher, and C.A. Mizzen, *Mass spectrometric characterization of human histone H3: a bird's eye view.* J Proteome Res, 2006. **5**(2): p. 240-7.

CHAPTER THREE

CONTINUOUS ELUTION PROTEOFORM IDENTIFICATION OF MYELIN BASIC PROTEIN BY SUPERFICIALLY POROUS REVERSED-PHASE LIQUID CHROMATOGRAPHY AND FOURIER TRANSFORM MASS SPECTROMETRY

This chapter was submitted for publication to a scientific journal in June, 2017 and is under review as of 7/24/17.

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 4 known MBP ASVs and hundreds of differentially modified states from a single 90 min SPLCMS run on ~0.5 μg of material. This included 323 proteoforms for 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 events linking palmitic acid, oleic acid and stearic acid at C78 of the 17.125 kDa ASV.

Introduction

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 (MScl) [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 function of the different ASVs is not well understood but their expression is dynamic with Golli-MBP ASVs expressed during embryonic development, followed by 17.12, 17.14, 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), and methionine sulfoxide (MSO), methionine sulfone (MSONE)), many of which are believed to regulate its structural and functional roles, such as influencing electrostatic interactions with the polyanionic species in the phospholipid bilayer and regulating signaling cascades through protein interactions [7, 8, 10-12]. Moscarello and colleagues have characterized PTMs 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, 11, 13-18]. Their work, and others, has begun to reveal intriguing insights on how MBP PTM status correlates with autoimmune activities in MScl and corresponding animal models (e.g., experimental autoimmune encephalomyelitis (EAE)) [19].

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 [1]. 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 [11]. 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 [20]. To address this need, this study reports on new top-down mass spectrometry (TDMS) [21] resources that prove effective for discrete proteoform characterization on MBP isolated from CNS tissues (i.e., brain and spine). TDMS typically employs both liquid chromatography, intact mass analysis, and gas-phase fragmentation (e.g., tandem mass spectrometry, MS/MS) to characterize a protein's sequence, localize the positions

of their PTMs, and quantify related proteoform ratios [22-24]. Herein we highlight the use of superficially porous reversed-phase liquid chromatography (SPLC), high-resolving power Fourier transform mass spectrometry (FTMS), and fragmentation by data-independent acquisition (DIA) with an aligned custom bioinformatics workflow [25, 26] that permits rapid interrogation of MBP. 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 acid (i.e, isobars). 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.

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 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 ºC. The extraction of MBP was accomplished via the Folch/acid protocol [27] as described in Appendix V.

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 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 $[M + 10H]^{10+}$ of ubiquitin by direct infusion. Samples were analyzed at a resolving power of 60,000 at 400 *m/z* and nozzle skimmer dissociation (NSD) potential of 25 V for intact mass studies, and NSD potential of 60 V and a capillary temperature of 375 °C for the data-independent SPLC-NSD-FTMS studies.

Database Generation and Data Analysis

A three-tiered absolute mass search strategy was employed for protein and proteoform identifications (Figure S-1, Appendix II). 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 annotated [28] MBP-ASV/PTM databases to differentiate the presence of unique MBP proteoforms. The third tier (Tier 3) was performed manually and used the Δm mode search in ProSightPC 3.0 (Thermo Fisher Scientific, Waltham, MA) to characterize unknown sequence variants and modifications [29].

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 min or 0.5 min. PD4 outputs were converted into observed mass, retention time, intensity and estimated grand average hydrophobicity index (GRAVY) values, the later determined by an internal calibration curve created from known proteins within the sample. Monoisotopic masses (denoted by the *-0* isotopologue label in figures) are consistently reported for intact proteoforms throughout the manuscript.

Database (DB) Generation: Tier 1 databases were compiled from 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 was captured MySQL Lite (XX) and theoretical isoelectric point and gravy data was calculated for each proteoform with ProPAS.1.1.pl [30]. 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 that is N-ACET and METH at positions 1 and 104, respectively. Tier 2 DBs for MBP were similarly generated from custom flat files specific to each ASV with each PTM considered as variable with the exception of fixed N-ACET (Supplemental Methods, Appendix II). PTMs and amino acid

abbreviations in text follow that described in the UniProt Knowledgebase (UniProtKB) [31]. The PTMs selected were based on previously reported PHOS, METH, N-ACET, and MSO/MSONE sites resulting from sequence alignment with multiple organisms [16]. Citrullination and deamidation were not examined in this study. For all MBP proteoforms discussed it is assumed each exists with fixed N-ACET.

Continuous elution absolute mass searches: Protein/proteoform identification was accomplished with a Visual Basic macros implemented in Excel (Microsoft…), Figure S-2, Appendix II) run on a Windows 2012 Server. The workflow supports generation of Poisson-based P-scores (*P*) of proteins/proteoforms [32] at sequential time intervals across an entire SPLC-NSD-FTMS dataset. The workflow also automates 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 dataset and the separate SPLC-FTMS dataset performed without NSD. DB searches started with an absolute mass test utilizing a 1.2 Dalton (Da) mass tolerance. Each candidate sequence is then tested against the fragment data at all time points simultaneously across the SPLC-NSD-FTMS run with a *P(i)* reported at every time point *i*. To avoid spurious matches across the entire time range, a P_{decoy} value was determined coinciding to the real target search at each *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 sought to ensure 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 $P_{decoy}(i)$ values were corrected for (divided by) the

number of related proteoforms in the database for the target protein sequence ($P'_{decoy}(i)$). Then a 99th percent confidence value at each i (denoted $P'_{decoy}(i),_{\alpha=0.01}$) 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 $P(i)$ - $P_{decoy}(i),_{\alpha=0.01}$ and only $P'(i) < 0.01$ were considered true hits (Figure S-3, Appendix II). 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 Validation: Along with ranking proteins/proteoform identifications by $P'(i)$, sequence maps were plotted for each identifications, with each hit automatically ranked from 0-5 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-4, Appendix II). 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. Final tabulated outputs for all protein/proteoform identifications are grouped by theoretical monoisotopic masses which presents proteoforms with similar chemical composition together. Heat maps generated reflects the relative -$Log_{10}(P')$ intensity for hits over time in regions 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) resins [33] with a capillary LC interfaced online with FTMS because early tests on different RPLC resin configurations provided evidence 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 μg of protein from brain tissue isolate were separated with a 90 minute linear gradient both without and with DIA by NSD. A representative total ion current (TIC) chromatogram, fragmentation, and protein feature maps (Figure 1) highlight the sample contains numerous protein/proteoforms with masses from 7-22 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 Appendix III). Compared to non-MBP proteins each of the MBP ASVs eluted relatively early in the SPLC runs 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 shows rodent MBP predominately exists as the 14.1 kDa and 18.4 kDa ASVs.

Identification of Classical MBP ASVs

Mass spectra corresponding to the peak elution period (centroid) for the 4 detected classical ASVs suggested three of four variants consisted of numerous proteoforms in varied relative abundance (Figure 2, left & middle). Amino acid sequence maps resulting from informatics searches on NSD data corresponding to the same elution period confirm the presence of 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 kDa, 18.4 kDa, and 17.125 kDa ASVs each exhibited inter-proteoform mass differences consistent with METH, Δ 14 Da, DIMETH, Δ 28 Da), and PHOS, Δ 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.

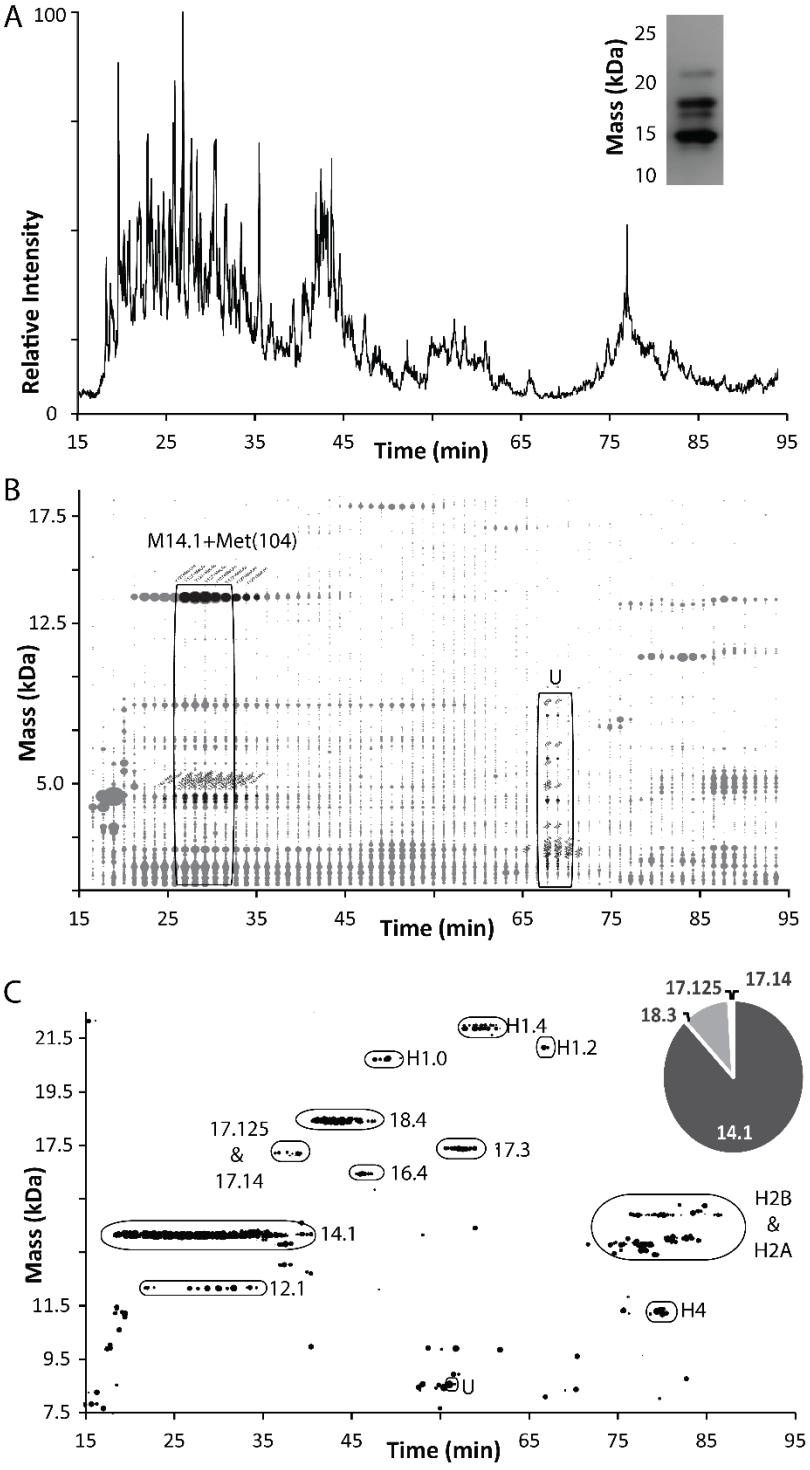

A
100
0
Relative Intensity
15
25
35
45
Time (min)
65
75
85
95
Mass (kDa)
25
20
15
10
B
17.5
12.5
5.0
Mass (kDa)
M14.1+Met(104)
U
15
25
35
45
Time (min)
65
75
85
95
C
21.5
19.5
17.5
11.5
9.5
7.5
Mass (kDa)
H1.4
H1.2
H1.0
18.4
17.125
&
17.14
17.3
16.4
14.1
12.1
H2B
&
H2A
H4
U
15
25
35
45
Time (min)
65
75
85
95
17.125
17.14
18.3
14.1

Figure 1

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 *y*-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 1A. Encircled are proteoform groups associated with the 12.1, 14.1, 16.4, 17.125, 17.14, 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).

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 1). A proteoform feature map (Figure 3A) and mass spectra of the $[M+15H]^{+15}$ charge state at six distinct time intervals (Figure 3B-G) highlight intensity fluctuations of 13 groups of apparent nominal mass isobars over time (labeled a-m) (Figure 3A, bar inset). Inspection of inter-proteoform mass differences (Figure 3B-3G, bar insets) relative to the N-ACET proteoform (a) is consistent with a mixture of MSO and MSONE early in the elution period (e.g., 3B-D) followed by elution of the N-ACET proteoform (Figure 3E), then the METH (Δ 14 Da, Figure 3F) and then DIMETH proteoforms (Δ 28 Da, Figure 3G). While PHOS (Δ 80 Da) without and with METH (Δ 94 Da) and DIMETH (Δ 108 Da) was generally observed throughout the elution period, DIPHOS proteoforms without and with METH and DIMETH (e.g., Δ 160 Da, Δ 174 Da, and Δ 188 Da) as well as TRIPHOS proteoforms (not shown) are 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-3G, insets). For example, the presence of a M19 and M124 MSO was observed early in the elution period on b81 fragment (left inset) and y43 fragment (right inset),

respectively. The y43 fragment also confirms the N-ACET proteoform elutes prior to the METH and then DIMETH proteoforms (Figure 3E-3F insets), highlighting that increased methylation content increases hydrophobicity of the 14 kDa ASV.

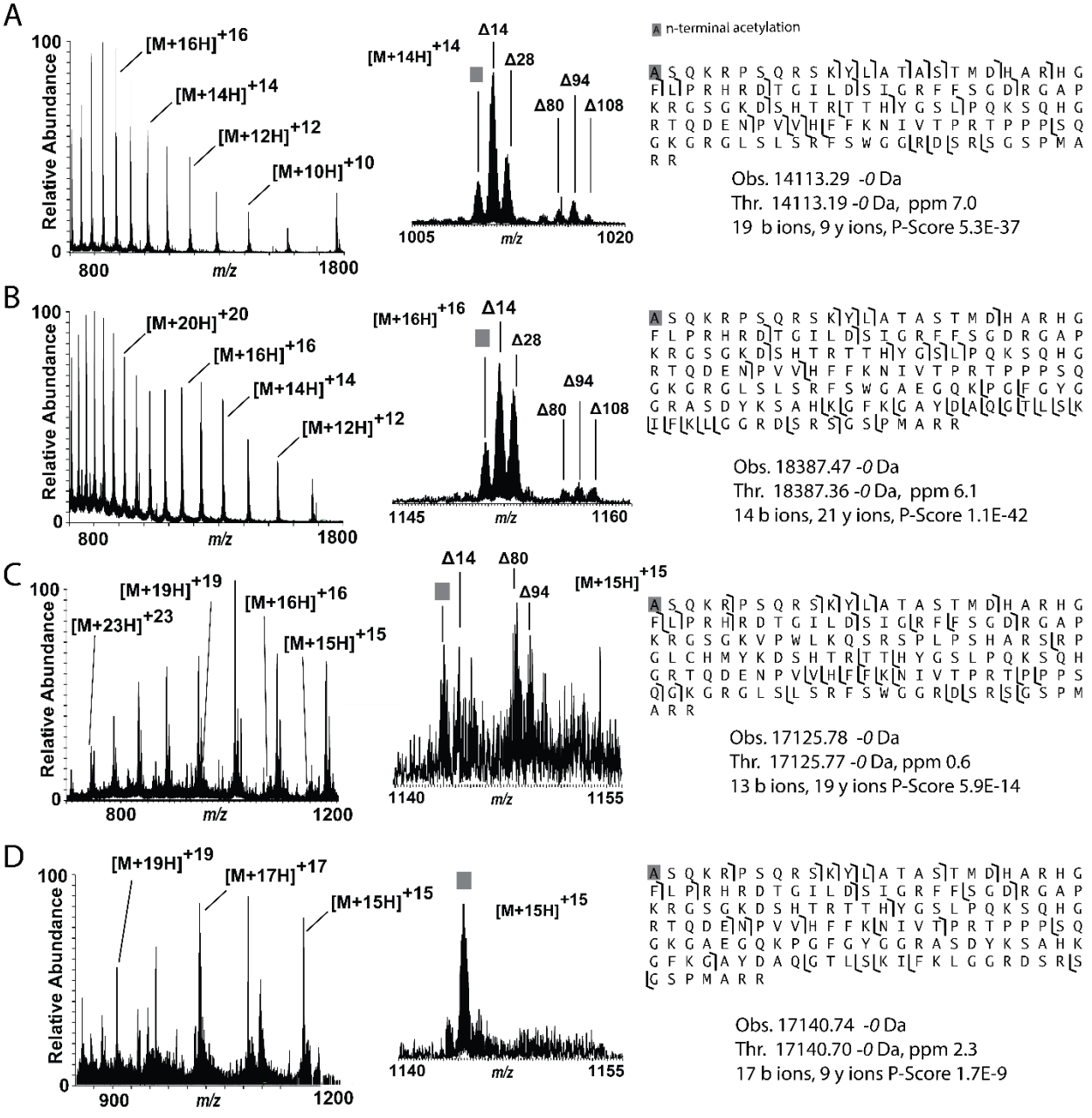


Figure 2

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 (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 a SPLC-NSD-FTMS run.

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 Appendix IV). 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 could be 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 2-4 with two 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 confirms 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 and 4C) 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 (14129 Da), MSO+METH (14143 Da), and MSONE (14145 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 and 3D) as well as corroborated

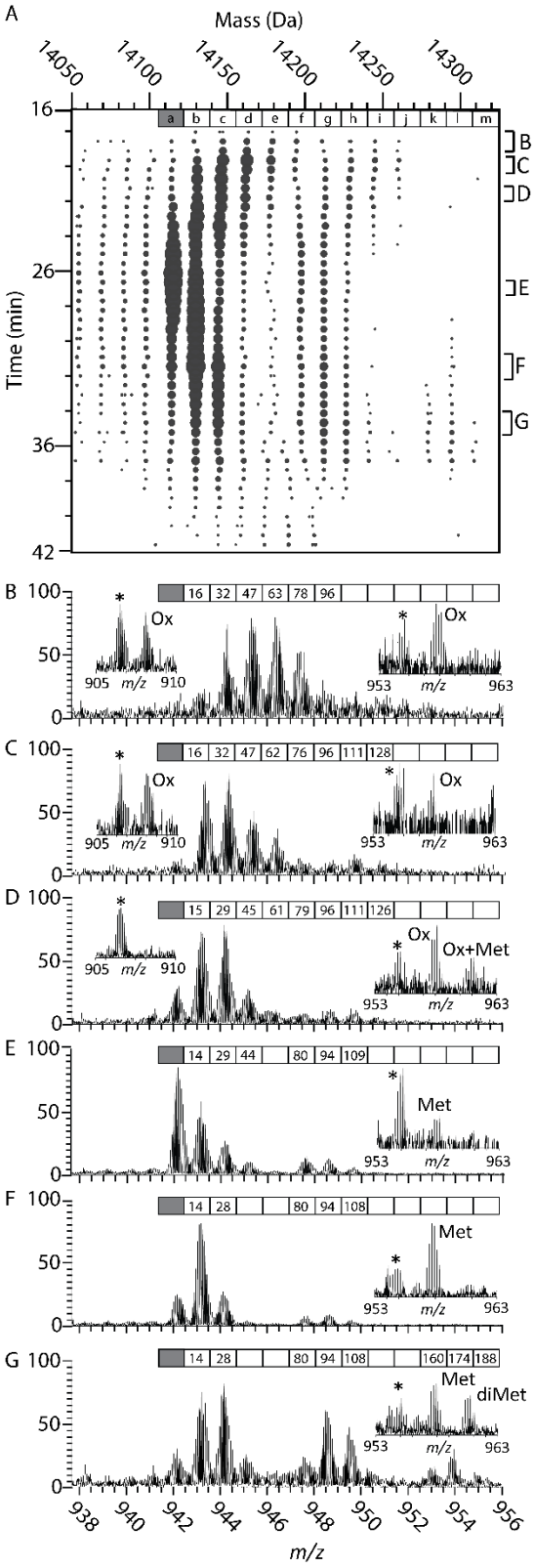

A
Mass (Da)
14050
14100
14150
14200
14250
14300
16
26
36
42
Time (min)
a b c d e f g h i j k l m
B
C
D
E
F
G
B 100
50
0
16 32 47 63 78 96
Ox
Ox
905 m/z 910
953 m/z 963
C 100
16 32 47 62 76 96 111 128
Ox
Ox
905 m/z 910
953 m/z 963
D 100
15 29 45 61 79 96 111 126
Ox
Ox+Met
905 m/z 910
953 m/z 963
E 100
14 29 44 80 94 109
Met
953 m/z 963
F 100
14 28 80 94 108
Met
953 m/z 963
G 100
14 28 80 94 108 160 174 188
Met
diMet
953 m/z 963
938 940 942 944 946 948 950 952 954 956
m/z

Figure 3

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 3A. Labels *a-m* in the bar inset (3A) designate apparent nominal mass proteoforms detected throughout the analysis. The bar insets (3B-3G) highlight common delta mass differences (e.g., Δ 14 Da, Δ 16 Da, Δ 80 Da, etc.) observed between spectral partners for groups *a-m* across the elution period. Mass spectra insets of the *b*81 ion (left) highlights elution of a MSO-M19 proteoform early in the run (3B-3C vs. 3D). Mass spectra insets of *y*43 ion (right) highlights the MSO-M124 proteoform prior to the METH-R104 and DIMET-R104 later in the run (3B-3G). Ox: MSO; Met: METH; diMet: DIMETH; *: unmodified fragment.

in separate replicate runs (not shown). The data also reveal that METH (14127 Da) and DIMETH (14141 Da) occurs 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) [34].

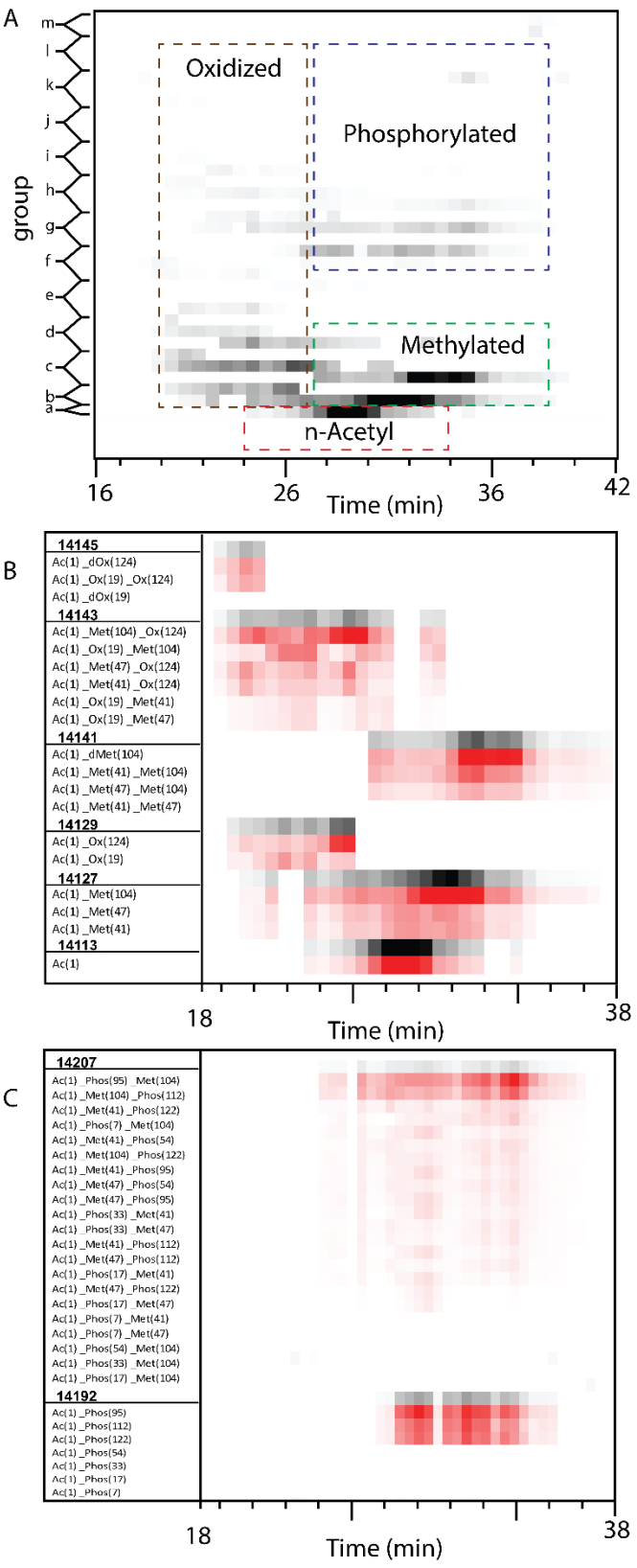

A
group
m
l
k
j
i
h
g
f
e
d
c
b
a
Oxidized
Phosphorylated
Methylated
n-Acetyl
16
26
36
42
Time (min)
B
14145
Ac(1) _dOx(124)
Ac(1) _Ox(19) _Ox(124)
Ac(1) _dOx(19)
14143
Ac(1) _Met(104) _Ox(124)
Ac(1) _Ox(19) _Met(104)
Ac(1) _Met(47) _Ox(124)
Ac(1) _Met(41) _Ox(124)
Ac(1) _Ox(19) _Met(41)
Ac(1) _Ox(19) _Met(47)
14141
Ac(1) _dMet(104)
Ac(1) _Met(41) _Met(104)
Ac(1) _Met(47) _Met(104)
Ac(1) _Met(41) _Met(47)
14129
Ac(1) _Ox(124)
Ac(1) _Ox(19)
14127
Ac(1) _Met(104)
Ac(1) _Met(47)
Ac(1) _Met(41)
14113
Ac(1)
18
38
Time (min)
C
14207
Ac(1) _Phos(95) _Met(104)
Ac(1) _Met(104) _Phos(112)
Ac(1) _Met(41) _Phos(122)
Ac(1) _Phos(7) _Met(104)
Ac(1) _Met(41) _Phos(54)
Ac(1) _Met(104) _Phos(122)
Ac(1) _Met(41) _Phos(95)
Ac(1) _Met(47) _Phos(54)
Ac(1) _Met(47) _Phos(95)
Ac(1) _Phos(33) _Met(41)
Ac(1) _Phos(33) _Met(47)
Ac(1) _Met(41) _Phos(112)
Ac(1) _Met(47) _Phos(112)
Ac(1) _Phos(17) _Met(41)
Ac(1) _Met(47) _Phos(122)
Ac(1) _Phos(17) _Met(47)
Ac(1) _Phos(7) _Met(41)
Ac(1) _Phos(7) _Met(47)
Ac(1) _Phos(54) _Met(104)
Ac(1) _Phos(33) _Met(104)
Ac(1) _Phos(17) _Met(104)
14192
Ac(1) _Phos(95)
Ac(1) _Phos(112)
Ac(1) _Phos(122)
Ac(1) _Phos(54)
Ac(1) _Phos(33)
Ac(1) _Phos(17)
Ac(1) _Phos(7)
18
38
Time (min)

Figure 4

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 Appendix IV). (B) Expanded heat map highlights the relative contributions of the unique proteoforms (red) N-ACET (14113 Da), METH (14127 Da), MSO (14129 Da), DIMETH (14141 Da), METH+MSO (14143 Da), and MSONE (14145 Da) to their respective average P' (black). (C) Expanded heat map for the numerous unique PHOS (14192 Da), and PHOS+METH (14207 Da) proteoforms. Ac: N-ACET, Met: METH, Ox: MSO, dMet: DIMETH, dOx: MSONE, Phos: PHOS.

Novel MBP ASVs

The initial absolute mass searches against the general mouse proteome revealed several components that did not immediately translate to confident identifications. For example, proteoform groups with masses of approximately 12.1 kDa and 16.4 kDa co-elute 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 and 5B vs. Fig 2A and 2B, middle). In both cases the lowest molecular mass proteoform for the 12.1 kDa and 16.4 kDa proteoform groups had a -1,961.04-0 Da mass difference relative to the 14.1 and 18.4 kDa ASVs, respectively. To localize the sequence discrepancy we searched NSD data using the Δm mode search feature 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 and 5B), suggesting both the 12.1 and 16.4 kDa variants derive from 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 [35], and exhibit methionine N-ACET [36].

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, mass spectrum showed proteoform microheterogeneity and ratio data consistent to that of the 14.1 and 18.4 kDa ASVs (Figure 5C). 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 5C). A Δ 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 5D). 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 [37].

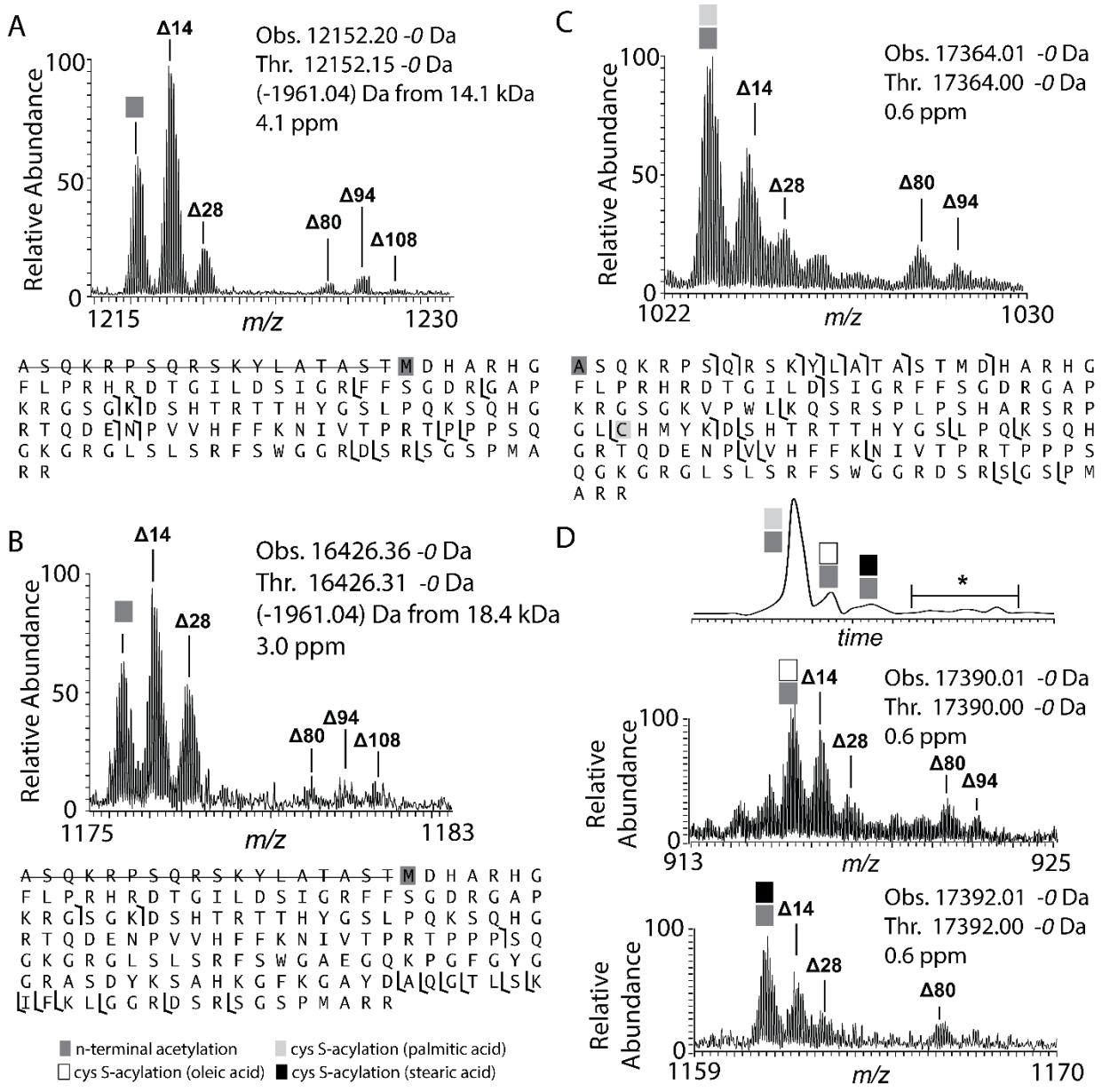


Figure 5

(A-B) Mass spectra associated with a single charge state at the peak elution period (centroid) of the novel 12.1 kDa and 16.4 kDa ASVs with corresponding fragment map confirming the identification. (C) Spectrum and fragment map confirms the addition of palmitic acid (16:0) to the 17.125 kDa ASV through S-acylation of C78 (1C). (D) Several proteoforms associated with the addition of oleic acid (18:2), stearic acid (18:0), as well as, other long chain fatty acids (*) to the 17.125 kDa ASV were chromatographically resolved and detected by accurate intact mass only.

Concluding Remarks

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 sought to develop a TDMS workflow that enables interrogation of MBP ASVs and abundant PTM states simultaneously. The described workflow shows significant potential at proteoform discovery with an information content superior to antibody or peptide based assays yet on a similar timescale 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 datasets. Within a single SPLC analysis we successfully identified 4 classical MBP ASVs and for the 14.1 kDa ASV we also 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. The current study was limited in that it employed DIA strategies that fragmented multiple proteoforms eluting concomitantly. While the bioinformatics resources enabled identification of the most abundant forms, a targeted approach using data-dependent CID, ETD, or HCD could yield more identifications and reduce ambiguity that can be caused by a fragment that can be assigned to multiple co-eluting proteoforms of MBP. Similarly, PTM neutral loss (e.g., phosphate or lipids) associated with the NSD method warrants prospective examination of labile PTMs via ECD or ETD. 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 [38]. It is plausible that both ASV expression and PTM expression of MBP are dynamic during the stages of development. Use of the platform in the 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 be correlated with the progression and remission of multiple sclerosis in humans.

References

1. Boggs, J.M., *Myelin basic protein: a multifunctional protein.* Cell Mol Life Sci, 2006. **63**(17): p. 1945-61.
2. Schmitt, S., L.C. Castelvetri, and M. Simons, *Metabolism and functions of lipids in myelin.* Biochim Biophys Acta, 2015. **1851**(8): p. 999-1005.
3. Poliak, S. and E. Peles, *The local differentiation of myelinated axons at nodes of Ranvier.* Nat Rev Neurosci, 2003. **4**(12): p. 968-80.
4. Love, S., *Demyelinating diseases.* J Clin Pathol, 2006. **59**(11): p. 1151-9.
5. Harauz, G. and D.S. Libich, *The classic basic protein of myelin--conserved structural motifs and the dynamic molecular barcode involved in membrane adhesion and protein-protein interactions.* Curr Protein Pept Sci, 2009. **10**(3): p. 196-215.
6. Vassall, K.A., V.V. Bamm, and G. Harauz, *MyelStones: the executive roles of myelin basic protein in myelin assembly and destabilization in multiple sclerosis.* Biochem J, 2015. **472**(1): p. 17-32.
7. Harauz, G. and J.M. Boggs, *Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms.* J Neurochem, 2013. **125**(3): p. 334-61.
8. Kim, J.K., et al., *Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis.* Mol Cell Proteomics, 2003. **2**(7): p. 453-62.
9. Akiyama, K., et al., *Study of expression of myelin basic proteins (MBPs) in developing rat brain using a novel antibody reacting with four major isoforms of MBP.* J Neurosci Res, 2002. **68**(1): p. 19-28.
10. Abbott, W.M., et al., *Soluble bovine galactose-binding lectin. cDNA cloning reveals the complete amino acid sequence and an antigenic relationship with the major encephalitogenic domain of myelin basic protein.* Biochem J, 1989. **259**(1): p. 283-90.
11. Harauz, G., et al., *Myelin basic protein-diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis.* Micron, 2004. **35**(7): p. 503-42.
12. Pritzker, L.B., et al., *Deimination of myelin basic protein. 2. Effect of methylation of MBP on its deimination by peptidylarginine deiminase.* Biochemistry, 2000. **39**(18): p. 5382-8.
13. Kim, J., et al., *Post-translational modifications of chicken myelin basic protein charge components.* Neurochem Res, 2009. **34**(2): p. 360-72.
14. Zand, R., et al., *Studies of posttranslational modifications in spiny dogfish myelin basic protein.* Neurochem Res, 2001. **26**(5): p. 539-47.
15. Martenson, R.E., M.J. Law, and G.E. Deibler, *Identification of multiple in vivo phosphorylation sites in rabbit myelin basic protein.* J Biol Chem, 1983. **258**(2): p. 930-7.
16. Zhang, C., et al., *Myelin basic protein undergoes a broader range of modifications in mammals than in lower vertebrates.* J Proteome Res, 2012. **11**(10): p. 4791-802.
17. Zand, R., et al., *Determination of the sites of posttranslational modifications in the charge isomers of bovine myelin basic protein by capillary electrophoresis-mass spectroscopy.* Biochemistry, 1998. **37**(8): p. 2441-9.
18. Vassall, K.A., et al., *Substitutions mimicking deimination and phosphorylation of 18.5-kDa myelin basic protein exert local structural effects that subtly influence its global folding.* Biochim Biophys Acta, 2016. **1858**(6): p. 1262-77.

19. Cao, L., D. Sun, and J.N. Whitaker, *Citrullinated myelin basic protein induces experimental autoimmune encephalomyelitis in Lewis rats through a diverse T cell repertoire.* J Neuroimmunol, 1998. **88**(1-2): p. 21-9.
20. Constantinescu, C.S., et al., *Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS).* Br J Pharmacol, 2011. **164**(4): p. 1079-106.
21. Patrie, S.M., *Top-Down Mass Spectrometry: Proteomics to Proteoforms.* Adv Exp Med Biol, 2016. **919**: p. 171-200.
22. Janssen, K.A., S. Sidoli, and B.A. Garcia, *Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology.* Methods Enzymol, 2017. **586**: p. 359-378.
23. Zhang, J., et al., *Proteoform analysis of lipocalin-type prostaglandin D-synthase from human cerebrospinal fluid by isoelectric focusing and superficially porous liquid chromatography with Fourier transform mass spectrometry.* Proteomics, 2014. **14**(10): p. 1223-31.
24. Zhang, J., et al., *Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography.* Anal Chem, 2013. **85**(21): p. 10377-84.
25. Patrie, S.M., et al., *Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa.* J Am Soc Mass Spectrom, 2004. **15**(7): p. 1099-108.
26. Patrie, S.M., et al., *Top down mass spectrometry of < 60-kDa proteins from Methanosarcina acetivorans using quadrupole FRMS with automated octopole collisionally activated dissociation.* Mol Cell Proteomics, 2006. **5**(1): p. 14-25.
27. Maatta, J.A., et al., *Detection of myelin basic protein isoforms by organic concentration.* Biochem Biophys Res Commun, 1997. **238**(2): p. 498-502.
28. Pesavento, J.J., et al., *Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry.* J Am Chem Soc, 2004. **126**(11): p. 3386-7.
29. Zamdborg, L., et al., *ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W701-6.
30. Wu, S. and Y. Zhu, *ProPAS: standalone software to analyze protein properties.* Bioinformation, 2012. **8**(3): p. 167-9.
31. Boutet, E., et al., *UniProtKB/Swiss-Prot.* Methods Mol Biol, 2007. **406**: p. 89-112.
32. Meng, F., et al., *Informatics and multiplexing of intact protein identification in bacteria and the archaea.* Nat Biotechnol, 2001. **19**(10): p. 952-7.
33. Roth, M.J., et al., *Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry.* Anal Chem, 2011. **83**(24): p. 9586-92.
34. Riley, N.M., M.S. Westphall, and J.J. Coon, *Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins.* Anal Chem, 2015. **87**(14): p. 7109-16.
35. Helbig, A.O., et al., *Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome.* Mol Cell Proteomics, 2010. **9**(5): p. 928-39.
36. Polevoda, B. and F. Sherman, *N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins.* J Mol Biol, 2003. **325**(4): p. 595-622.

37. Ross, N.W. and P.E. Braun, *Acylation in vitro of the myelin proteolipid protein and comparison with acylation in vivo: acylation of a cysteine occurs nonenzymatically.* J Neurosci Res, 1988. **21**(1): p. 35-44.
38. Bjelke, B. and A. Seiger, *Morphological distribution of MBP-like immunoreactivity in the brain during development.* Int J Dev Neurosci, 1989. **7**(2): p. 145-64.

CHAPTER FOUR

FUTURE DIRECTIONS

Future work using the platform of superficially porous reversed phase liquid chromatography mass spectrometry and continuous elution proteoform scoring could be used in a number of studies to better understand neurodegenerative processes as well as normal neurological development. The rodent EAE model can be employed in a number of contexts to investigate different aspects of neurodegenerative pathobiology. High resolution spatial mapping may be possible as the yield of MBP extract enables characterization of MBP from regions as small as 1-2 mm^3 which could enable characterization of proteoforms of MBP on individual lesions. Experiments during normal neurological development may characterize changes in MBP during development and provide a point of comparison between MBP in neurodegenerative disease states (which has been shown to be developmentally immature) and MBP from that has been isolated from young animals [1]. The maturation of MBP in the rodent brain has been shown to proceed in a rostral-caudal direction in the ventral spinal cord, and in both rostral and caudal directions in the dorsal spinal cord [2]. It has also been shown the MBP matures earlier in the cerebellum and spinal cord than cerebrum [3]. MBP extracted from different spatial regions at different times to identify the proteoforms that are present during different stages of neurological development and aid in the understanding of myelination, demyelination, and remyelination processes. These areas could also be combined together synergistically (i.e., comparing MBP from lesions from different modes of EAE induction that give rise to different neurodegenerative disease courses, or comparison of MBP from different clinical-subtypes of multiple sclerosis in human patients). The three-tiered workflow that has been introduced also enables the informatics procedures associated with the platform to be executed on an LC timescale, reducing the

bottleneck that once prevented studies comparing complex biological states. The three-tiered search procedure also confidently identifies common proteoforms leaving more time for the user to interrogate unique proteoforms like the proteoforms from populations at 12.1, 16.4, or 17.3 kDa in the previous study. It is possible that novel PTMs and splice variants will be identified and associated with demyelination, remyelination, or axonal loss.

Induction of experimental autoimmune encephalomyelitis (EAE) in rodent models and comparison of MBP from EAE and wild type rodents forms the basis for the most logical comparison that will inform outstanding issues related to multiple sclerosis and neurodegeneration and recovery that were previously discussed, most notably understanding the role of MBP in the different clinical subtypes of multiple sclerosis, and the opportunity to define biomarkers that could define subtype-specific therapy using specific proteoforms of myelin basic protein [4, 5]. Studies that include a comparison of MBP isolated from human tissue between different clinical subtypes (relapsing-remitting, secondary-progressive, primary-progressive, and progressive-relapsing) could complement animal work that induces specific subtypes of disease in different models of neurodegeneration. Future work could use a combination of spinal cord homogenate induction in SJL mice to model acute, monophasic disease for comparison to wild type SJL tissue. Induction using PLP139-151 in SJL mice could model relapsing-remitting disease course and allow comparison to wildtype SJL mice. Chronic disease could be modeled in C57BL/6 Mice induced with MOG33-55 and compared to wild type C57BL/6 mice [6].

References

1. Moscarello, M.A., et al., *Myelin in multiple sclerosis is developmentally immature.* J Clin Invest, 1994. **94**(1): p. 146-54.
2. Foran, D.R. and A.C. Peterson, *Myelin acquisition in the central nervous system of the mouse revealed by an MBP-Lac Z transgene.* J Neurosci, 1992. **12**(12): p. 4890-7.
3. Akiyama, K., et al., *Study of expression of myelin basic proteins (MBPs) in developing rat brain using a novel antibody reacting with four major isoforms of MBP.* J Neurosci Res, 2002. **68**(1): p. 19-28.
4. Lassmann, H., *Experimental models of multiple sclerosis.* Rev Neurol (Paris), 2007. **163**(6-7): p. 651-5.
5. Lassmann, H., W. Bruck, and C. Lucchinetti, *Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy.* Trends Mol Med, 2001. **7**(3): p. 115-21.
6. Furlan, R., C. Cuomo, and G. Martino, *Animal models of multiple sclerosis.* Methods Mol Biol, 2009. **549**: p. 157-73.

APPENDIX I SUPPLEMENTAL INFORMATION FOR CHAPTER TWO

Supporting Information

Preparation of Mouse Heart Protein

Hearts were excised from anesthetized 12 week-old male C57/Blk6 mice (Harlan) and ventricles were snap frozen in liquid nitrogen. Each ventricle was ground into frozen heart powder using a pre-chilled porcelain mortar and pestle, and then stored at -80 ºC in ~15 mg (wet weight) aliquots. Total heart protein was prepared by thawing while simultaneously precipitating protein from an aliquot of frozen heart powder in 10% trichloroacetic acid/10 mM dithiothreitol (DTT). The precipitated protein fraction was washed free of acid with ethyl ether. For analysis by MS, precipitated protein was suspended in 2 M Urea, 300 mM NaCl, 10 mM DTT, 20 mM EDTA, and vigorously agitated with zirconium oxide beads using a Bullet Blender (Innovative Solutions, Beverly Hills, MI). Prior to SPLC/MS analysis, a fraction corresponding to ~1 mg total heart (wet weight) was desalted using a C4 Zip Tip (Millipore, Billerica, MA). Briefly, Zip Tips were wetted with 40% acetonitrile, 10% isopropanol, 0.3% formic acid, and 0.025% trifluoroacetic acid (TFA) in water and equilibrated with 5% acetonitrile, 0.3% formic acid, and 0.025% TFA in water. Equilibrated tips were loaded with 10 μL of protein suspension and washed 10 times with 10 μL 5% acetonitrile, 0.3% formic acid, and 0.025% (TFA) in water. Proteins were eluted by rinsing 10 times with 62% acetonitrile, 15% isopropanol, 0.3% formic acid, and 0.025% TFA in water. The desalted protein solution was dried down to 10 μL for LC/MS analysis. Mouse hearts were collected using procedures in accordance with guidelines instituted by the Animal Care and Use Committee at the University of Texas Southwestern Medical Center.

Preparation of HeLa Acid Extracted Proteins

Anchorage dependent HeLa S3 cells were grown to ~80% confluence at 37 ºC in 5% CO_2 on 75 cm^2 culture flasks using RPM1 1640 growth media containing 1% penicillin-streptomycin and 10% fetal bovine serum. Cells were released from the flask using 0.1% trypsin for 5 minutes and centrifuged at 1000 RCF for 5 minutes. The precipitated cells were washed twice with fetal bovine serum and pelleted at 1000 RCF. Cell pellets were then washed with phosphate buffered saline (PBS) and aliquoted into either 1 x 10^4 or 5 x 10^4 cell pellets and snap frozen for storage. To determine cell pellet quantity, cell pellets were resuspended in equal volumes PBS and Trypan blue and counted using a Bright-Line hemacytometer (Hausser Scientific, Horsham, PA). For acid extraction, cell pellets were thawed on ice and treated with 40 µL ice-cold 0.4 N H_2SO_4 with vortexing and shearing. The mixture was placed on ice for 60 minutes and centrifuged at 16000 RCF for 5 minutes. The supernate was removed and a 10 µL aliquot desalted using a C4 Zip Tip (*v.s.*).

LTQ-Orbitrap MS settings

For analysis of standard proteins the instruments were operated with settings: full MS mode, scan range from 800-2000 *m/z*, 60000 FTMS resolution (at 400 *m/z*), 2 microscans, and a maximum ion accumulation time of 100 ms (target ion count 5E6). For analysis of complex endogenous mixtures, the instruments were operated with settings: SIM MS mode, a scan range from 1100-1400 *m/z*, 60000 FTMS resolution (at 400 *m/z*), 3 microscans, and a 300 ms maximum ion accumulation time. Scanning in SIM mode provided improved S/N and reduced background signal for polymeric contaminants at low *m/z*. Data were processed using the Xtract algorithm within Xcalibur to generate decharged, deisotoped chromatograms. Peak widths and

heights were determined by plotting the mass range containing the monoisotopic protein mass (+/- 2 Da) for each component observed in the monoisotopic mass chromatograms generated by Xtract for the complex endogenous mixtures.

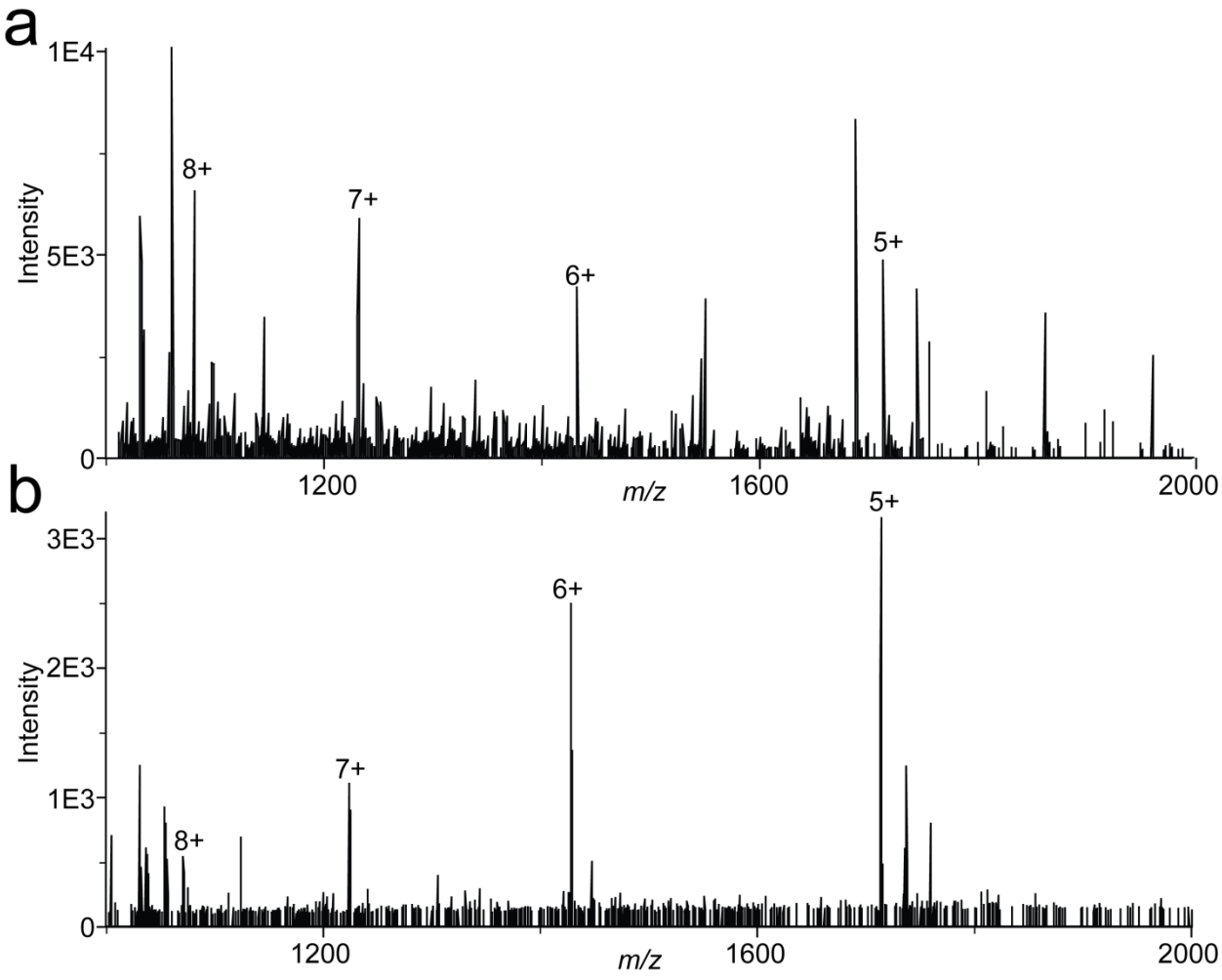


Figure S-1. Comparison of mass spectra at the LLOQ for ubiquitin on two different instruments. a) 100 amol on a LTQ Orbitrap XL and b) 20 amol on a LTQ Orbitrap Velos.

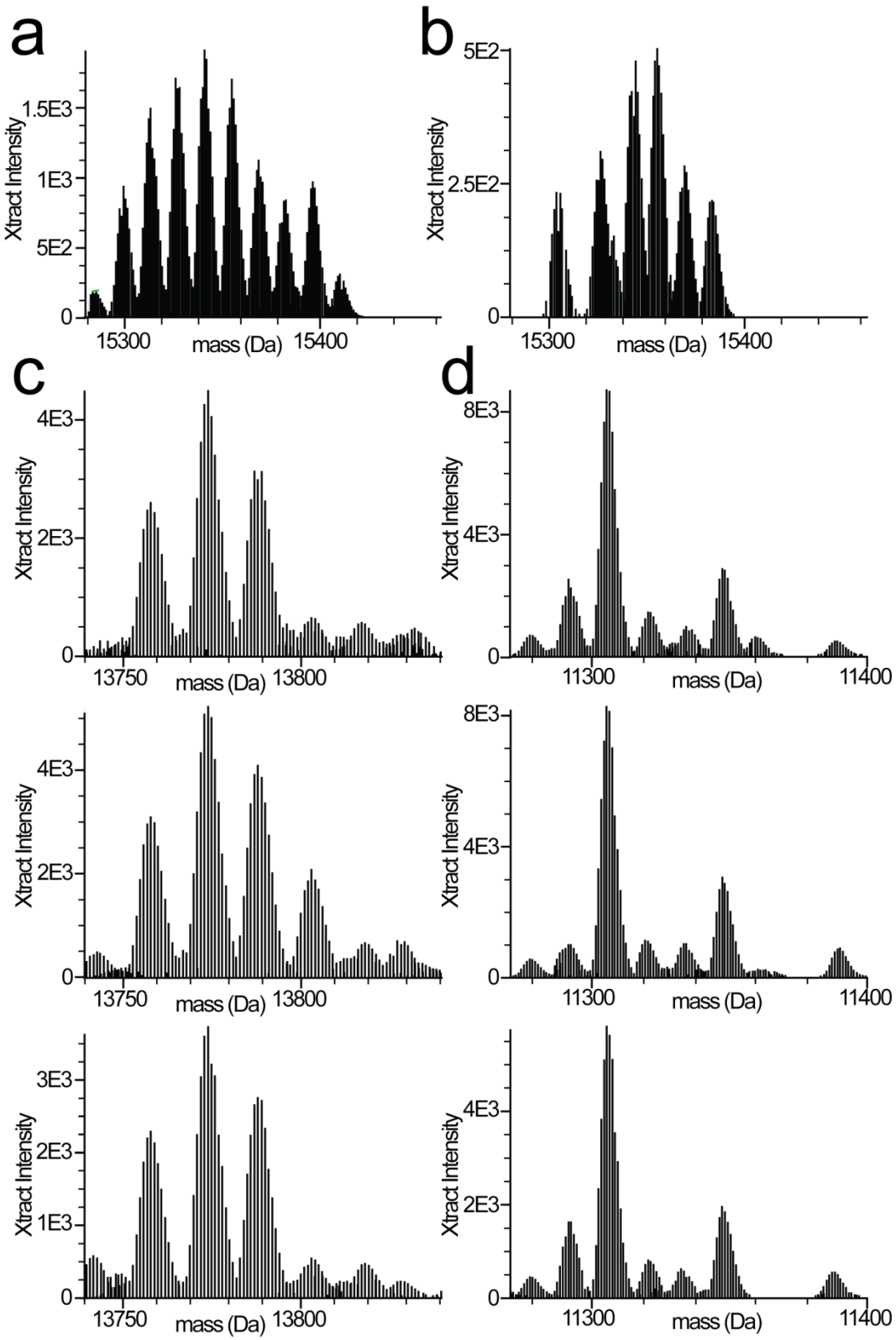

a
b
c
d
Xtract Intensity
mass (Da)
15300
15400
13750
13800
11300
11400
1.5E3
1E3
5E2
0
5E2
2.5E2
4E3
2E3
8E3
4E3
3E3
2E3
1E3

Figure S-2

Figure S-2. Evaluation of histone classes observed in SPLC/MS of HeLa acid extract. a) Xtract degcharged mass spectrum for histone H3.1 PTM profile from 5 x 10^4 cells and b) from 1 x 10^4 cells. c) Xtract decharged mass spectra for histone H2B protein gene isoforms and (d) histone H4 PTM isoforms from 1 x 10^4 cell replicates demonstrate small changes in absolute intensity, although relative signal intensities are consistent between the three runs.

	LTQ-Orbitrap XL			LTQ-Orbitrap Velos		
Protein	**S/N at 1 fmol**	**Calculated LOD (amol)**	**Calculated LLOQ (amol)**	**S/N at 1 fmol**	**Calculated LOD (amol)**	**Calculated LLOQ (amol)**
Ubiquitin*	87.3	34.4	114.5	243.3	12.3	41.1
Chicken Egg Lysozyme	5.6	535.7	1785.7	19.6	153.5	511.5
Bovine Serum Albumin	32.8	91.5	305.2	N/A	N/A	N/A
Horse Heart Myoglobin	10.3	290.5	968.3	122.1	24.6	81.9
Carbonic Anhydrase**	21.9	136.7	455.7	29.4	101.9	339.7

* Carbonic anhydrase standard contains ~10% ubiquitin contamination, ubiquitin LOD not corrected for this factor
** Carbonic anhydrase standard contains ~10% ubiquitin contamination and 15.6 kDa species (~10%), likely superoxide dismutase, carbonic anhydrase LOD not corrected for these factors

Table S-1 signal-to-noise and detection limits calculated from spectra generated at 1 fmol infjected on column for 5 proteins

Table S-2. Advantages and disadvantages of five common RPLC stationary phases for intact protein separations.

Stationary Phase	Advantages	Disadvantages
Porous Silica	High sample loading capacity, high dynamic range at "conventional" back pressures (<400 bar)	Significant mass transfer effects, limited speed of analyses, limited pH range (pH 2-8)
Nonporous silica	Minimal mass transfer effects, operation at high linear velocities using UPLC	Low loading capacity, small particles require high back pressures (>400 bar), limited pH range (pH 2-8)
Polymer (PS/DVB)	Minimal mass transfer effects, operation at high linear velocities using UPLC, large pH range (pH 1-13)	Low loading capacity, small particles require high back pressures (>400 bar)
Monolithic	Minimal mass transfer effects, operation at high linear velocities, large pH range (pH 1-13) for polymer monoliths	Limited loading capacity and dynamic range
SP Silica	Minimal mass transfer effects, operation at high linear velocities at <400 bar, high sample loading capacities, high dynamic range	Limited operating pH range (pH 2-8)

Table S-3. Variation observed for 25 replicate injections of ubiquitin at 1.75× LLOQ.

Parameter	Mean	RSD
EIC Peak Area	3.40E5	19%
EIC Peak Width (FWHM)	3.5 sec	20%
EIC Peak Retention Time	5.77 min	0.54%
Deisotoped Xtract Height	6.43E3	40%
MS Peak Height– $[M+5H^+]^{5+}$	4.2E3	21%
LOD – $[M+5H^+]^{5+}$	26.4 amol	67%

APPENDIX II SUPPLEMENTAL INFORMATION FOR CHAPTER THREE

Supplemental Methods

Flat file information used for custom highly annotated proteoform databases for MBP alternative splice variants (ASVs), included the following: (A) P04370-2, (B) P04370-4, (C) P04370-5, (D) P04370-6, (E) P04370-7, (F) P04370-8, (G) P04370-9, (H) P04370-10, (I) P04370-11.

Databases were generated in ProsightPC 3.0. Custom flat files were generated to incorporate PTMs that had been reported in previous literature or predicted by PTM prediction software. The ptm_info, ptm_type, and resid_swissprot databases in ProsightPC 3.0 were edited to include methionine sulfoxide. In the database creation wizard, methionine cleavage and n-terminal formylation were not introduced. N-terminal acetylation was introduced. SNPs were not introduced. The maximum number of features per sequence was 25, and the maximum mass was 70000. All PTMs were included. The 12152, 15164, 15179, and 16426 Da sequences were the same sequence as the 14113, 17152, 17140 and 18367 kDa splice variants, respectively, with an alternative transcriptional start site resulting in the protein beginning at methionine 19 of the classical sequences. Proteoforms that were not n-terminally acetylated were removed from the database.

(A)

```
ID   MBP_MOUSE               Reviewed;         195 AA.
AC   P04370-2;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    195       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES     151    151       Phosphoserine.
FT   MOD_RES     153    153       Methionine sulfone; alternate.
FT   MOD_RES     153    153       Methionine sulfoxide; alternate.
FT   MOD_RES     167    167       Phosphothreonine.
FT   MOD_RES     188    188       Phosphoserine.
SQ   SEQUENCE   195 AA;  20991 MW;  B418ED11C27B0C43 CRC64;
     MGNHSGKREL SAEKASKDGE IHRGEAGKKR SVGKLSQTAS EDSDVFGEAD AIQNNGTSAE
     DTAVTDSKHT ADPKNNWQGA HPADPGNRPH LIRLFSRDAP GREDNTFKDR PSESDELQTI
     QEDPTAASGG LDVMASQKRP SQRSKYLATA STMDHARHGF LPRHRDTGIL DSIGRFFSGD
     RGAPKRGSGK VSSEP
```

(B)

```
ID   MBP_MOUSE               Reviewed;         194 AA.
AC   P04370-4;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
```

```
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    194       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES     130    130       Omega-N-methylarginine; alternate.
FT   MOD_RES     130    130       Symmetric dimethylarginine; alternate.
FT   MOD_RES     145    145       N6-acetyllysine.
FT   MOD_RES     159    159       Phosphoserine.
FT   MOD_RES     173    173       Phosphothreonine.
FT   MOD_RES     189    189       Phosphoserine.
FT   MOD_RES     191    191       Methionine sulfone; alternate.
FT   MOD_RES     191    191       Methionine sulfoxide; alternate.
SQ   SEQUENCE   194 AA;  21357 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKVPWL
     KQSRSPLPSH ARSRPGLCHM YKDSHTRTTH YGSLPQKSQH GRTQDENPVV HFFKNIVTPR
     TPPPSQGKGR GLSLSRFSWG AEGQKPGFGY GGRASDYKSA HKGFKGAYDA QGTLSKIFKL
     GGRDSRSGSP MARR
```

(C)

```
ID   MBP_MOUSE               Reviewed;         168 AA.
AC   P04370-5;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    168       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES      95     95       Phosphothreonine.
FT   MOD_RES     104    104       Omega-N-methylarginine; alternate.
FT   MOD_RES     104    104       Symmetric dimethylarginine; alternate.
FT   MOD_RES     112    112       Phosphoserine.
FT   MOD_RES     119    119       N6-acetyllysine.
FT   MOD_RES     133    133       Phosphoserine.
FT   MOD_RES     147    147       Phosphothreonine.
FT   MOD_RES     163    163       Phosphoserine.
FT   MOD_RES     165    165       Methionine sulfone; alternate.
FT   MOD_RES     165    165       Methionine sulfoxide; alternate.
SQ   SEQUENCE   168 AA;  18345 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKDSHT
```

```
     RTTHYGSLPQ KSQHGRTQDE NPVVHFFKNI VTPRTPPPSQ GKGRGLSLSR FSWGAEGQKP
     GFGYGGRASD YKSAHKGFKG AYDAQGTLSK IFKLGGRDSR SGSPMARR
```

(D)

```
ID   MBP_MOUSE               Reviewed;         153 AA.
AC   P04370-6;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    153       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES      78     78       S-palmitoyl cysteine.
FT   MOD_RES     130    130       Omega-N-methylarginine; alternate.
FT   MOD_RES     130    130       Symmetric dimethylarginine; alternate.
FT   MOD_RES     138    138       Phosphoserine.
FT   MOD_RES     148    148       Phosphoserine.
FT   MOD_RES     150    150       Methionine sulfone; alternate.
FT   MOD_RES     150    150       Methionine sulfoxide; alternate.
SQ   SEQUENCE   153 AA;  17083 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKVPWL
     KQSRSPLPSH ARSRPGLCHM YKDSHTRTTH YGSLPQKSQH GRTQDENPVV HFFKNIVTPR
     TPPPSQGKGR GLSLSRFSWG GRDSRSGSPM ARR
```

(E)

```
ID   MBP_MOUSE               Reviewed;         157 AA.
AC   P04370-7;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    157       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES      95     95       Phosphothreonine.
FT   MOD_RES     108    108       N6-acetyllysine.
FT   MOD_RES     122    122       Phosphoserine.
FT   MOD_RES     136    136       Phosphothreonine.
FT   MOD_RES     152    152       Phosphoserine.
```

```
FT   MOD_RES     154    154       Methionine sulfone; alternate.
FT   MOD_RES     154    154       Methionine sulfoxide; alternate.
SQ   SEQUENCE   157 AA;  17098 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKDSHT
     RTTHYGSLPQ KSQHGRTQDE NPVVHFFKNI VTPRTPPPSQ GKGAEGQKPG FGYGGRASDY
     KSAHKGFKGA YDAQGTLSKI FKLGGRDSRS GSPMARR
```

(F)

```
ID   MBP_MOUSE               Reviewed;         127 AA.
AC   P04370-8;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    127       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES      95     95       Phosphothreonine.
FT   MOD_RES     104    104       Omega-N-methylarginine; alternate.
FT   MOD_RES     104    104       Symmetric dimethylarginine; alternate.
FT   MOD_RES     112    112       Phosphoserine.
FT   MOD_RES     122    122       Phosphoserine.
FT   MOD_RES     124    124       Methionine sulfone; alternate.
FT   MOD_RES     124    124       Methionine sulfoxide; alternate.
SQ   SEQUENCE   127 AA;  14071 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKDSHT
     RTTHYGSLPQ KSQHGRTQDE NPVVHFFKNI VTPRTPPPSQ GKGRGLSLSR FSWGGRDSRS
     GSPMARR
```

(G)

```
ID   MBP_MOUSE               Reviewed;         183 AA.
AC   P04370-9;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    183       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES     121    121       Phosphothreonine.
```

```
FT   MOD_RES     134    134       N6-acetyllysine.
FT   MOD_RES     148    148       Phosphoserine.
FT   MOD_RES     162    162       Phosphothreonine.
FT   MOD_RES     178    178       Phosphoserine.
FT   MOD_RES     180    180       Methionine sulfone; alternate.
FT   MOD_RES     180    180       Methionine sulfoxide; alternate.
SQ   SEQUENCE   183 AA;  20111 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIG RFFSGDRGAP KRGSGKVPWL
     KQSRSPLPSH ARSRPGLCHM YKDSHTRTTH YGSLPQKSQH GRTQDENPVV HFFKNIVTPR
     TPPPSQGKGA EGQKPGFGYG GRASDYKSAH KGFKGAYDAQ GTLSKIFKLG GRDSRSGSPM
     ARR
```

(H)

```
ID   MBP_MOUSE               Reviewed;         190 AA.
AC   P04370-10;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
FT   CHAIN         1    190       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES      95     95       Phosphothreonine.
FT   MOD_RES     126    126       Omega-N-methylarginine; alternate.
FT   MOD_RES     126    126       Symmetric dimethylarginine; alternate.
FT   MOD_RES     134    134       Phosphoserine.
FT   MOD_RES     141    141       N6-acetyllysine.
FT   MOD_RES     155    155       Phosphoserine.
FT   MOD_RES     169    169       Phosphothreonine.
FT   MOD_RES     185    185       Phosphoserine.
FT   MOD_RES     187    187       Methionine sulfone; alternate.
FT   MOD_RES     187    187       Methionine sulfoxide; alternate.
SQ   SEQUENCE   190 AA;  20670 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIGR FFSGDRGAP KRGSGKDSHT
     RTTHYGSLPQ KSQHGRTQDE NPVVHFFKNI VTPRTPPPSQG KDFVPGDHH VNVSVVTVSF
     SSSQGRGLSL SRFSWGAEGQ KPGFGYGGRA SDYKSAHKGFK GAYDAQGTL SKIFKLGGRD
     SRSGSPMARR
```

(I)

```
ID   MBP_MOUSE               Reviewed;         175 AA.
AC   P04370-11;
DT   20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT   18-OCT-2001, sequence version 2.
DT   22-JAN-2014, entry version 152.
DE   RecName: Full=Myelin basic protein;
DE            Short=MBP;
DE   AltName: Full=Myelin A1 protein;
GN   Name=Mbp; Synonyms=Shi;
OS   Mus musculus (Mouse).
```

```
FT   CHAIN         1    175       Myelin basic protein.
FT                                /FTId=PRO_0000158991.
FT   MOD_RES       7      7       Phosphoserine.
FT   MOD_RES      17     17       Phosphoserine.
FT   MOD_RES      19     19       Methionine sulfone; alternate.
FT   MOD_RES      19     19       Methionine sulfoxide; alternate.
FT   MOD_RES      33     33       Phosphothreonine.
FT   MOD_RES      54     54       Phosphoserine.
FT   MOD_RES     121    121       Phosphothreonine.
FT   MOD_RES     152    152       Omega-N-methylarginine; alternate.
FT   MOD_RES     152    152       Symmetric dimethylarginine; alternate.
FT   MOD_RES     159    159       Phosphoserine.
FT   MOD_RES     170    170       Phosphoserine.
FT   MOD_RES     172    172       Methionine sulfone; alternate.
FT   MOD_RES     172    172       Methionine sulfoxide; alternate.
SQ   SEQUENCE   175 AA;  19408 MW;  B418ED11C27B0C43 CRC64;
     ASQKRPSQRS KYLATASTMD HARHGFLPRH RDTGILDSIGR FFSGDRGAP KRGSGKVPWL
     KQSRSPLPSH ARSRPGLCHM YKDSHTRTTH YGSLPQKSQHG RTQDENPVV HFFKNIVTPR
     TPPPSQGKDF VPGDHHVNVS VVTVSFSSSQ GRGLSLSRFSW GGRDSRSGS PMARR
//
```

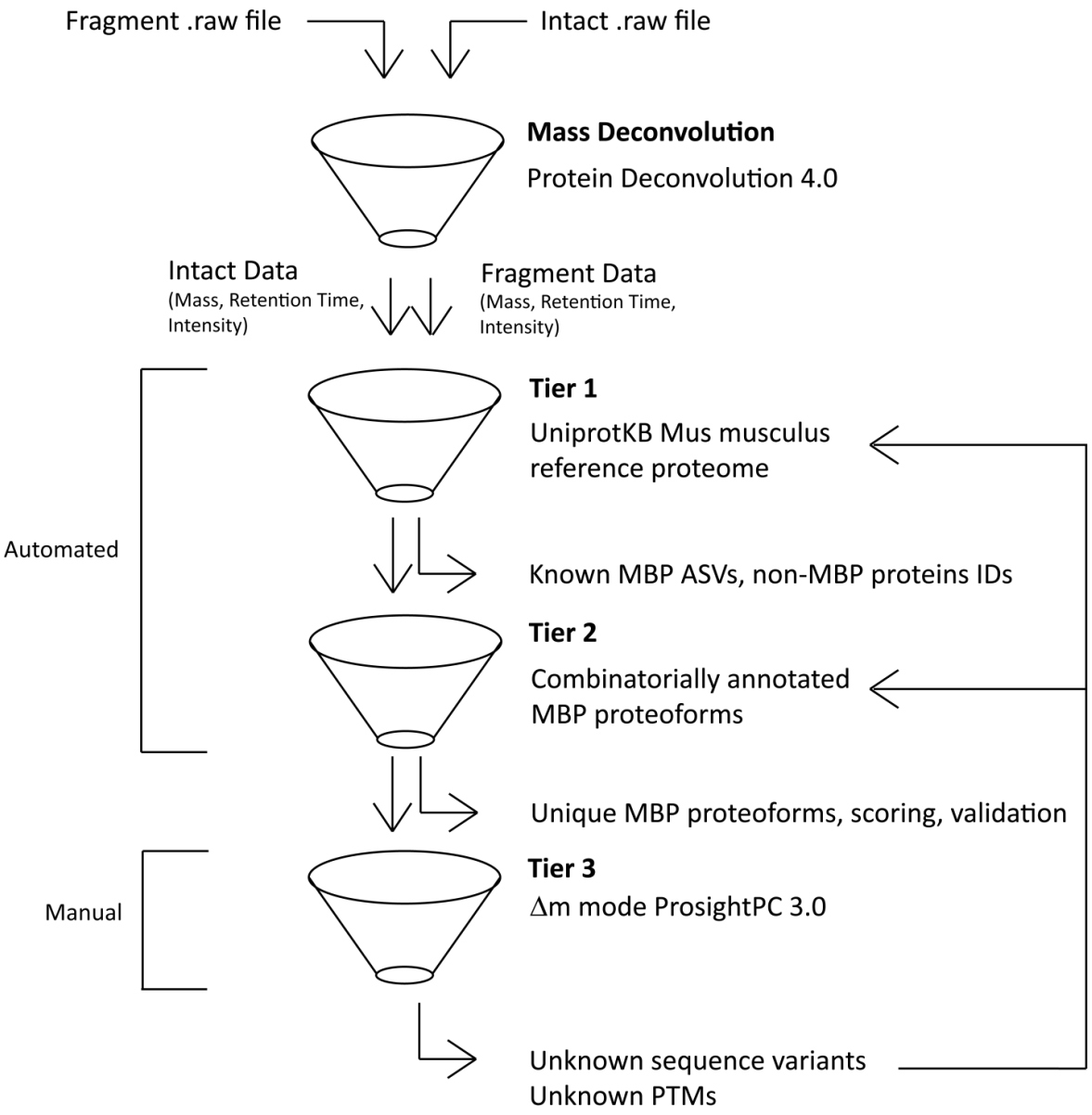


Figure S-1: Three-tiered workflow. Tier 1 is automated and identifies major classical splice variants of MBP and non-MBP proteins. Tier 2 is automated and identifies proteoforms from combinatorially annotated databases for each splice variant. Tier 3 manually employs the Δm Mode from ProsightPC 3.0 to identify novel splice variants and PTMs which are later incorporated into Tier 1 or Tier 2 searches.

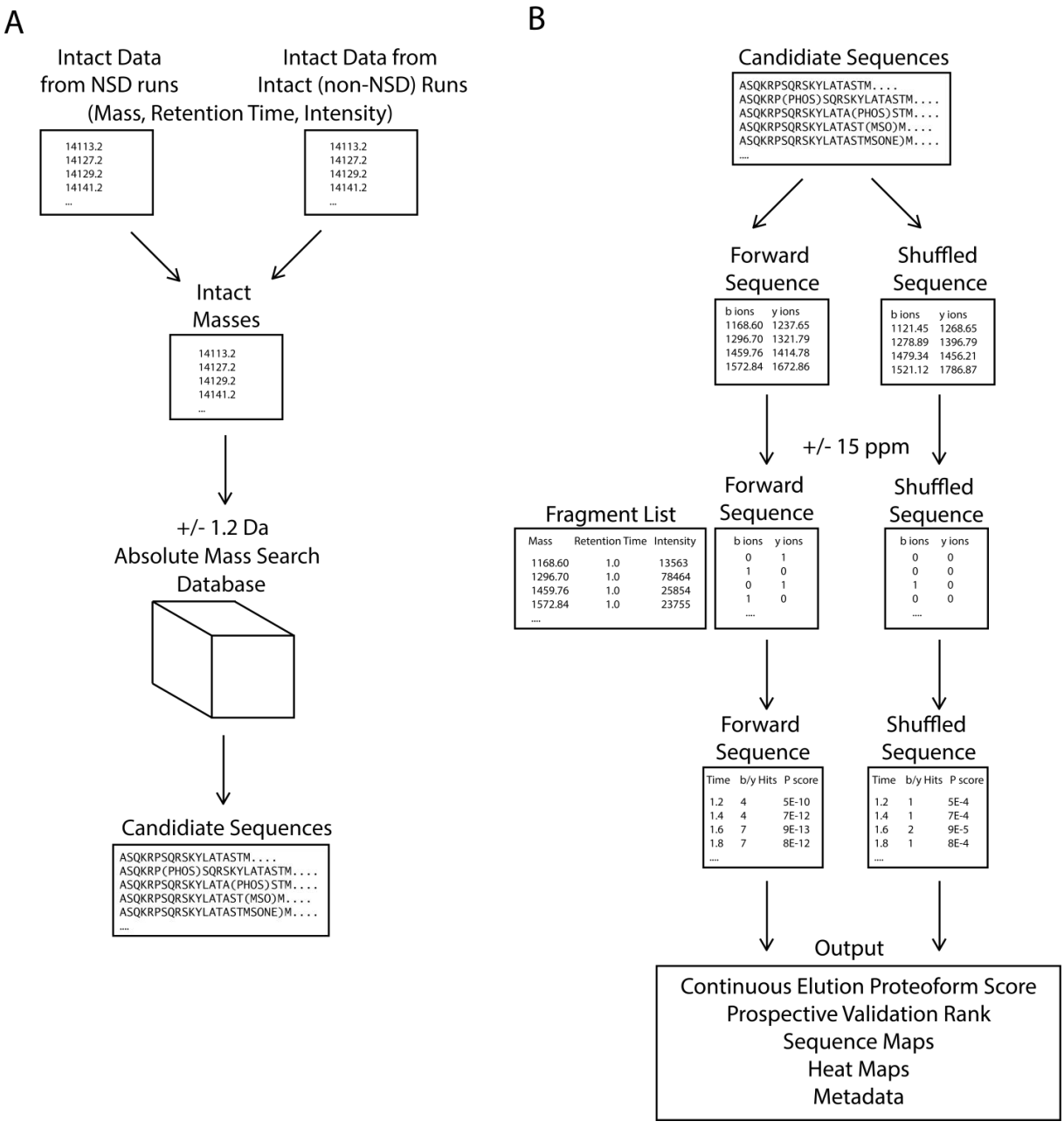

A
Intact Data from NSD runs (Mass, Retention Time, Intensity)
14113.2
14127.2
14129.2
14141.2
...
Intact Data from Intact (non-NSD) Runs
14113.2
14127.2
14129.2
14141.2
...
Intact Masses
14113.2
14127.2
14129.2
14141.2
...
+/- 1.2 Da Absolute Mass Search Database
Candidiate Sequences
ASQKRPSQRSKYLATASTM....
ASQKRP(PHOS)SQRSKYLATASTM....
ASQKRPSQRSKYLATA(PHOS)STM....
ASQKRPSQRSKYLATAST(MSO)M....
ASQKRPSQRSKYLATASTMSONE)M....
....
B
Candidiate Sequences
ASQKRPSQRSKYLATASTM....
ASQKRP(PHOS)SQRSKYLATASTM....
ASQKRPSQRSKYLATA(PHOS)STM....
ASQKRPSQRSKYLATAST(MSO)M....
ASQKRPSQRSKYLATASTMSONE)M....
....
Forward Sequence
b ions y ions
1168.60 1237.65
1296.70 1321.79
1459.76 1414.78
1572.84 1672.86
Shuffled Sequence
b ions y ions
1121.45 1268.65
1278.89 1396.79
1479.34 1456.21
1521.12 1786.87
+/- 15 ppm
Fragment List
Mass Retention Time Intensity
1168.60 1.0 13563
1296.70 1.0 78464
1459.76 1.0 25854
1572.84 1.0 23755
....
Forward Sequence
b ions y ions
0 1
1 0
0 1
1 0
....
Shuffled Sequence
b ions y ions
0 0
0 0
1 0
0 0
....
Forward Sequence
Time b/y Hits P score
1.2 4 5E-10
1.4 4 7E-12
1.6 7 9E-13
1.8 7 8E-12
....
Shuffled Sequence
Time b/y Hits P score
1.2 1 5E-4
1.4 1 7E-4
1.6 2 9E-5
1.8 1 8E-4
....
Output
Continuous Elution Proteoform Score
Prospective Validation Rank
Sequence Maps
Heat Maps
Metadata

Figure S-2

Figure S-2: Overview of the data processing search schema implemented. (A) Raw (.raw) data files from intact (non-NSD) runs or NSD runs were processed with Protein Deconvolution 4.0 to yield intact masses and fragment masses. These were searched against a full proteome (Tier 1) or MBP specific (Tier 2) database with a 1.2 Da tolerance to yield candidate sequences. (B) For Tier 2 searches, fragment ions were searched against candidate (forward) sequences as well as shuffled decoy sequences. A fragment list was generated where b and y ions present at each retention time were recorded. The fragment lists at each retention time were compiled and used in the adjusted p value calculation to yield the continuous elution proteoform score. The prospective validation rank, sequence maps, heat maps, and metadata were also generated.

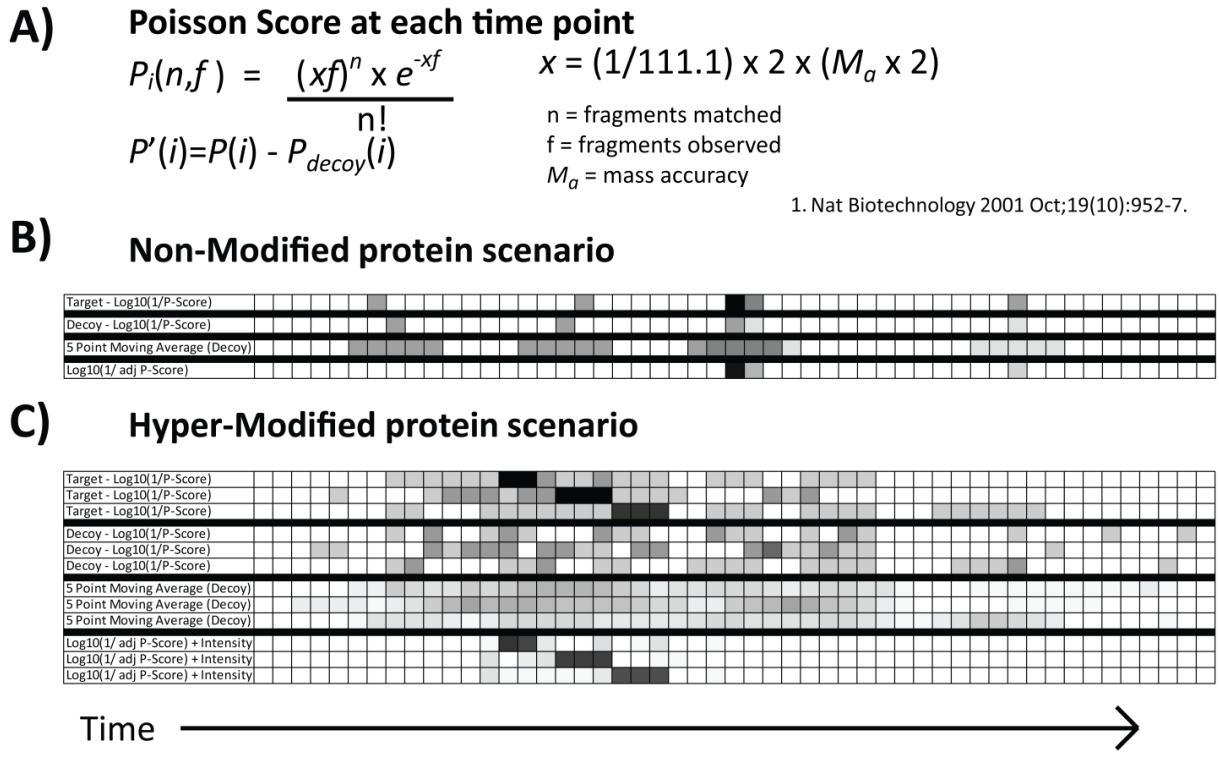


Figure S-3: Schema highlights essential components of continuous elution scoring and heat map generation. (A) The candidate sequences were tested against the fragment data at all time points for all IEF fractions simultaneously with a baseline-adjusted Poisson-based p-score (*P*') reported for any hits at each time point (i). $P'(i) = P(i) - P_{decoy}(i)$; where *P* is calculated as described by Meng et. al. [1] and $P_{decoy}(i)$ is determined simultaneously on an inverted and scrambled sequence that retains the candidate's PTMs at the same amino acid location to that of the target. Generally, the baseline P_{decoy} value at each (*i*) was determined at a 95% confidence interval after the individual decoy scores were smoothed by 5 point moving average. (B) A heat map containing continuous elution scoring for *P*, P_{decoy}, and *P*' for a single protein. (C) A heat map containing continuous elution scoring for *P*, P_{decoy}, and *P*' for multiple proteoforms with high sequence / PTM identity. The calculation of *P*'(*i*) included the requirement that the intact mass of the sequence must be observed in the input data.

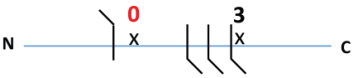
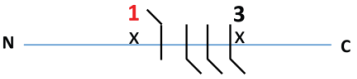
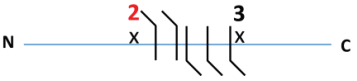
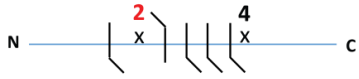
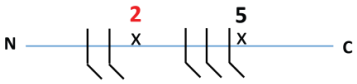
Rank	Rule Description	Example	Lowest Count (per PTM)	Count (all PTMs)*	Rank Descriptor
0	0 Fragment on at least 1 PTM + ~~0 Fragments/all PTMs~~		0	N/A	Not Confirmed
1	1 Fragment on each PTM + ~~0 Fragments/all PTMs~~		1	N/A	Plausible
2	= 2 Fragments/PTM + 0 Fragments/all PTMs		2	0	Valid
3	= 2 Fragments/PTM + 1 Fragments/all PTMs		2	1	Valid
4	= 2 Fragments/PTM + =2 Fragments/all PTMs		2	2	Valid
5	Protein with no PTMs		N/A	N/A	Valid

Figure S-4: Validation of proteoforms by ranking. A rank of 0 indicates 0 fragments validating at least 1 internal PTM. A rank of 1 indicates at least 1 fragment validating each internal PTM. A rank of 2 indicates at least 2 fragments validating each internal PTM and 0 fragments validating all internal PTMs. A rank of 3 indicates at least 2 fragments validating each internal PTM and 1 fragment validating all internal PTMs. A rank of 4 indicates at least 2 fragments per internal PTM and at least 2 fragments validating all internal PTMs. A rank of 5 indicates validation of a protein without PTMs.

APPENDIX III NON-MBP PROTEOFORM IDENTIFICATIONS

Table S-1 Non-MBP proteoform identifications

Description	Obs Mass	Thr Mass	Mass Error (Da)	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14117.9 Average										
Histone H2A type 1	14117.92	14117.9	0.03	2.12	20	3	23	1.55E-14	**valid**	2	2	4
Histone H2A type 1	14117.92	14117.9	0.03	2.12	20	3	23	4.79E-10	**valid**	2	0	2
Histone H2A type 1	14117.92	14117.9	-0.0064	-0.45	2	16	18	1.82E-03	**valid**	2	1	3
Histone H2A type 1	14117.92	14117.9	0.03	2.12	1	14	15	7.76E-10	plausible	1	1	1
		14080 Average										
Histone H2A type 1	14079.9	14080	-0.0687	-4.88	5	14	19	6.31E-10	**valid**	2	0	2
Histone H2A type 1	14079.9	14080	-0.0687	-4.88	5	14	19	1.29E-09	**valid**	2	0	2
		14079.9 Average										
Histone H2A type 1	14079.9	14079.9	-0.0323	-2.29	7	13	20	6.17E-10	**valid**	7	7	4
		14078 Average										
Histone H2A type 1-F	14077.88	14078	-0.072	-5.11	16	3	19	2.95E-12	**valid**	2	2	4
		14077.9 Average										
Histone H2A type 2-A	14077.88	14077.9	0.0168	1.19	2	14	16	3.02E-08	plausible	1	1	1
		14077.8 Average										
Histone H2A type 2-A	14077.88	14077.8	0.0532	3.78	20	3	23	2.82E-15	**valid**	2	2	4
Histone H2A type 2-A	14077.88	14077.8	0.0532	3.78	20	3	23	4.17E-14	**valid**	2	0	2
		14066 Average										
Histone H2A type 1	14064.93	14066	-1.019	-72.44	6	14	20	5.75E-15	**valid**	6	6	4
Histone H2A type 1	14064.93	14066	-1.019	-72.44	6	14	20	3.63E-11	**valid**	6	6	4
Histone H2A type 3	14064.93	14066	-1.019	-72.44	3	17	20	4.17E-19	**valid**	3	0	2
		14063.9 Average										
Histone H2A type 1-F	14063.96	14063.9	0.0256	1.82	20	4	24	1.66E-15	**valid**	20	20	4
		14063.8 Average										
Histone H2A type 2-A	14063.96	14063.8	0.1145	8.14	1	15	16	1.48E-10	plausible	1	1	1
		14052 Average										
Histone H2A type 1-K	14051.98	14052	0.0072	0.51	8	3	11	2.09E-04	**valid**	7	6	4
		14051.9 Average										
Histone H2A type 1	14051.98	14051.9	0.0436	3.1	18	16	34	2.29E-39	**valid**	3	3	4
Histone H2A type 1	14051.98	14051.9	0.0436	3.1	18	16	34	1.23E-31	**valid**	3	1	3
Histone H2A type 1-K	14051.98	14051.9	0.0436	3.1	23	4	27	6.17E-22	**valid**	23	23	4
Histone H2A type 3	14051.98	14051.9	0.0436	3.1	4	15	19	8.51E-11	**valid**	4	4	4
		14039.9 Average										
Histone H2A type 2-A	14039.87	14039.9	-0.003	-0.21	6	15	21	6.03E-11	**valid**	6	6	4
Histone H2A type 2-A	14039.87	14039.9	-0.0394	-2.81	2	16	18	8.13E-17	**valid**	2	0	2
Histone H2A type 2-A	14039.87	14039.9	-0.0394	-2.81	2	16	18	2.00E-16	**valid**	2	0	2
		14038 Average										
Histone H2A type 1	14037.96	14038	0.0069	0.49	8	16	24	3.89E-23	**valid**	8	6	4
		14037.9 Average										
Histone H2A type 1	14037.96	14037.9	0.0433	3.08	23	15	38	2.00E-46	**valid**	23	23	4
Histone H2A type 1	14037.96	14037.9	0.0433	3.08	23	15	38	1.74E-44	**valid**	23	23	4
Histone H2A type 3	14037.96	14037.9	0.0433	3.08	8	16	24	3.02E-21	**valid**	8	8	4
		14012.4 Average										
Histone H2B type 3-B	14012.91	14012.4	0.4592	32.77	2	6	8	1.07E-03	plausible	1	0	1
Histone H2B type 3-B	14012.91	14012.4	0.4592	32.77	2	6	8	1.23E-03	plausible	1	0	1
		14011.9 Average										
Histone H2A type 2-A	14012.91	14011.9	1.0339	73.79	6	10	16	2.63E-15	**valid**	2	2	4
Histone H2A type 2-A	14012.91	14011.9	1.0339	73.79	6	10	16	1.38E-10	**valid**	2	1	3

Description	Obs Mass	Thr Mass	Mass Error (Da)	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		13997.9 Average										
Histone H2A type 2-A	13997.92	13997.9	0.0586	4.19	20	15	35	5.75E-47	**valid**	20	20	4
Histone H2A type 2-A	13997.92	13997.9	0.0586	4.19	20	15	35	2.29E-43	**valid**	20	20	4
Histone H2A type 2-A	13997.92	13997.9	0.0222	1.59	9	16	25	7.76E-24	**valid**	7	5	4
		13947.9 Average										
Histone H2A.J	13947.95	13947.9	0.0309	2.22	2	5	7	1.70E-03	**valid**	2	2	4
		13946.8 Average										
Histone H2A type 1-H	13947	13946.8	0.2107	15.11	12	2	14	3.98E-11	**valid**	2	2	4
		13890.8 Average										
Histone H2A type 2-C	13889.89	13890.8	-0.9191	-66.17	20	3	23	7.41E-20	**valid**	20	20	4
Histone H2A type 2-C	13889.89	13890.8	-0.9191	-66.17	20	3	23	2.88E-15	**valid**	20	20	4
Histone H2A type 2-C	13889.89	13890.8	-0.9555	-68.79	9	2	11	8.91E-07	**valid**	5	5	4
		13890.5 Average										
Histone H2B type 3-B	13889.89	13890.5	-0.581	-41.83	3	9	12	6.92E-04	**valid**	3	3	4
		13852.8 Average										
Histone H2A type 1-H	13853.56	13852.8	0.7597	54.84	17	5	22	1.48E-10	**valid**	17	17	4
		13852.6 Average										
Histone H2B type 1-P	13853.56	13852.6	0.973	70.24	0	9	9	6.31E-05	**No PTMs**	9	0	5
		13852.5 Average										
Histone H2B type 1-B	13853.56	13852.5	1.0934	78.93	1	6	7	1.51E-03	plausible	1	1	1
		13828.4 Average										
Histone H2B type 1-F/J/L	13827.57	13828.4	-0.7921	-57.28	4	5	9	5.75E-07	plausible	1	0	1
		13820.5 Average										
Histone H2B type 1-H	13821.55	13820.5	1.0703	77.44	2	6	8	8.13E-05	plausible	1	1	1
Histone H2B type 1-H	13821.55	13820.5	1.0703	77.44	1	6	7	1.86E-04	plausible	1	2	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	6.03E-04	plausible	1	1	1
Histone H2B type 1-H	13821.55	13820.5	1.0703	77.44	1	6	7	3.55E-04	plausible	1	2	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	1.17E-03	plausible	1	1	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	1.45E-03	plausible	1	1	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	8.91E-04	plausible	1	1	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	1.12E-03	plausible	1	1	1
Histone H2B type 1-H	13821.55	13820.5	1.0703	77.44	1	6	7	5.50E-04	plausible	1	2	1
Histone H2B type 1-H	13821.55	13820.5	1.0703	77.44	1	6	7	6.61E-04	plausible	1	2	1
Histone H2B type 2-B	13821.55	13820.5	1.0703	77.44	1	5	6	8.71E-04	plausible	1	1	1
		13813.5 Average										
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	6.92E-05	plausible	1	1	1
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	7.08E-05	plausible	1	1	1
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	4.57E-04	plausible	1	1	1
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	4.57E-04	plausible	1	1	1
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	5.25E-04	plausible	1	1	1
Histone H2B type 3-B	13812.56	13813.5	-0.8922	-64.59	1	6	7	1.58E-03	plausible	1	1	1
		13812.4 Average										
Histone H2B type 2-B	13812.56	13812.4	0.1899	13.75	4	6	10	4.68E-04	**valid**	2	0	2
		13811.5 Average										
Histone H2B type 1-C/E/G	13811.57	13811.5	0.1001	7.25	1	6	7	3.80E-05	plausible	1	1	1

Description	Obs Mass	Thr Mass	Mass Error (Da)	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
Histone H2B type 3-B	13810.55	13810.5	0.0454	3.29	1	10	11	3.09E-07	plausible	1	1	1
Histone H2B type 3-B	13810.55	13810.5	0.0454	3.29	1	10	11	1.41E-06	plausible	1	1	1
Histone H2B type 3-B	13810.55	13810.5	0.0454	3.29	1	10	11	2.19E-04	plausible	1	1	1
Histone H2B type 3-B	13810.55	13810.5	0.0454	3.29	1	10	11	2.00E-04	plausible	1	1	1
Histone H2B type 1-B	13810.55	13810.5	0.093	6.73	0	6	6	4.68E-04	plausible	1	1	1
Histone H2B type 1-B	13810.55	13810.5	0.093	6.73	0	6	6	1.05E-03	plausible	1	1	1
Histone H2B type 3-A	13810.55	13810.5	0.0818	5.92	0	6	6	1.29E-03	plausible	1	1	1
Histone H2B type 1-B	13810.55	13810.5	0.093	6.73	0	6	6	3.02E-03	plausible	1	1	1
		13796.5 Average										
Histone H2B type 1-M	13796.58	13796.5	0.0526	3.81	0	18	18	7.41E-12	**No PTMs**	18	0	5
Histone H2B type 1-F/J/L	13796.58	13796.5	0.0526	3.81	1	12	13	6.76E-04	**No PTMs**	13	0	5
		13794.5 Average										
Histone H2B type 1-F/J/L	13795.56	13794.5	1.0929	79.23	2	6	8	3.89E-06	plausible	1	1	1
Histone H2B type 1-F/J/L	13795.56	13794.5	1.0929	79.23	3	6	9	3.24E-06	plausible	1	1	1
Histone H2B type 1-F/J/L	13795.56	13794.5	1.0929	79.23	2	6	8	5.75E-06	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	1.95E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	1.17E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	2.34E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	1.58E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	0	6	6	3.47E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	3.63E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	1.45E-03	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	8.13E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	8.13E-04	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	2.88E-03	plausible	1	1	1
Histone H2B type 1-M	13795.56	13794.5	1.0929	79.23	1	5	6	2.29E-04	plausible	1	1	1
		13780.5 Average										
Histone H2B type 1-H	13780.57	13780.5	0.0416	3.02	2	17	19	1.38E-20	**No PTMs**	19	0	5
Histone H2B type 2-B	13780.57	13780.5	0.0416	3.02	1	18	19	9.33E-21	**No PTMs**	19	0	5
		13778.5 Average										
Histone H2B type 1-H	13779.57	13778.5	1.0968	79.6	1	6	7	2.24E-04	plausible	1	1	1
Histone H2B type 2-B	13779.57	13778.5	1.0968	79.6	1	5	6	1.32E-03	plausible	1	1	1
Histone H2B type 1-H	13779.57	13778.5	1.0968	79.6	0	6	6	9.12E-04	plausible	1	1	1
		13766.5 Average										
Histone H2B type 1-C/E/G	13766.56	13766.5	0.0442	3.21	1	18	19	7.59E-22	**No PTMs**	19	0	5
Histone H2B type 1-P	13766.56	13766.5	0.0554	4.02	1	7	8	2.82E-06	plausible	1	1	1
Histone H2B type 1-P	13766.56	13766.5	0.0554	4.02	1	7	8	1.66E-04	plausible	1	1	1
Histone H2B type 1-P	13766.56	13766.5	0.0554	4.02	1	7	8	1.48E-03	plausible	1	1	1
		13764.4 Average										
Histone H2B type 1-B	13764.52	13764.4	0.1214	8.82	1	6	7	2.34E-04	plausible	1	1	1
		13748.4 Average										
Histone H2B type 1-M	13748.55	13748.4	0.1523	11.08	2	5	7	1.48E-03	**valid**	2	2	4
		13736.5 Average										
Histone H2B type 1-H	13736.48	13736.5	0.0174	1.27	3	5	8	2.69E-06	plausible	1	1	1
Histone H2B type 1-H	13736.48	13736.5	0.0174	1.27	3	5	8	4.07E-05	plausible	1	1	1
Histone H2B type 2-B	13736.48	13736.5	0.0174	1.27	3	5	8	5.89E-05	plausible	1	1	1
Histone H2B type 1-H	13736.48	13736.5	0.0174	1.27	3	5	8	4.37E-05	plausible	1	1	1
Histone H2B type 2-B	13736.48	13736.5	0.0174	1.27	1	5	6	1.48E-03	plausible	1	1	1
Histone H2B type 2-B	13736.48	13736.5	0.0174	1.27	1	5	6	1.15E-03	plausible	1	1	1
		13720.4 Average										
Histone H2B type 3-B	13720.49	13720.4	0.1186	8.64	2	6	8	1.41E-04	**valid**	2	2	4

Description	Obs Mass	Thr Mass	Mass Error (Da)	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		11355.4 Average										
Histone H4	11355.44	11355.4	0.0673	5.93	0	9	9	2.45E-05	plausible	1	1	1
		11341.4 Average										
Histone H4	11341.42	11341.4	0.0226	1.99	1	8	9	9.33E-08	plausible	1	1	1
Histone H4	11341.42	11341.4	0.0226	1.99	1	8	9	1.17E-04	plausible	1	1	1
Histone H4	11341.42	11341.4	0.0226	1.99	1	8	9	2.88E-04	plausible	1	1	1
		11299.4 Average										
Histone H4	11299.4	11299.4	0.0211	1.87	1	8	9	1.45E-05	plausible	1	1	1
Histone H4	11299.4	11299.4	0.0211	1.87	1	8	9	4.68E-05	plausible	1	1	1
Histone H4	11299.4	11299.4	0.0211	1.87	2	8	10	2.00E-06	plausible	1	1	1
Histone H4	11299.4	11299.4	0.0211	1.87	1	8	9	4.79E-04	plausible	1	1	1
		8559.6 Average										
Ubiquitin	8559.665	8559.6	0.0484	5.65	5	11	16	1.86E-16	**No PTMs**	16	0	5

APPENDIX IV MBP PROTEOFORM IDENTIFICATIONS FOR THE 14.1 KDA SPLICE VARIANT

Summary	
Unique protein identifiers →	**1**
Unique proteoforms →	**323**
Hits "Rank 1" →	**252**
Hits "Rank 2-4" →	**71**
Hits "No PTMs" (*rank 5) →	**0**

Table S-1 Summary of MBP proteoform identifications for the 14.1 kDa splice variant

Table S-2 MBP Proteoform identifications for the 14.1 kDa splice variant

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14303.1 Average									
P04370-8_Ac(1) _Phos(17) _Ox(19) _Phos(33) _Met(47)	14303.2	14303	4.68	9	4	13	5.37E-08	**valid**	2	2	4
P04370-8_Ac(1) _Ox(19) _Phos(33) _Phos(54) _Met(104)	14303.2	14303	4.68	9	15	24	1E-21	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Phos(54) _Met(104)	14303.2	14303	4.68	8	15	23	3.89E-18	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Phos(33) _Met(104)	14303.2	14303	4.68	8	15	23	1.7E-12	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Phos(54) _Met(104)	14303.2	14303	4.68	1	15	16	4.68E-12	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Phos(17) _Ox(19) _Met(104)	14303.2	14303	4.68	1	15	16	2.29E-11	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Phos(95) _Met(104)	14303.2	14303	4.68	10	5	15	6.17E-09	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Phos(33) _Met(104)	14303.2	14303	4.68	1	15	16	4.47E-10	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(54) _Phos(95) _Met(104)	14303.2	14303	4.68	11	5	16	4.07E-07	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Phos(95) _Met(104)	14303.2	14303	4.68	9	5	14	5.01E-06	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(47) _Phos(54)	14303.2	14303	4.68	9	4	13	6.46E-07	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Phos(54)	14303.2	14303	4.68	9	4	13	3.98E-07	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(54) _Phos(95)	14303.2	14303	4.68	11	3	14	2.69E-07	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Phos(95)	14303.2	14303	4.68	10	3	13	1.62E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Phos(54)	14303.2	14303	4.68	8	4	12	0.00012	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(54) _Phos(95)	14303.2	14303	4.68	10	3	13	2.57E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Phos(95)	14303.2	14303	4.68	9	3	12	1.86E-06	plausible	1	1	1
		14301.2 Average									
P04370-8_Ac(1) _Phos(33) _Met(47) _Phos(54) _Met(104)	14302.2	14301	70.28	10	16	26	6.46E-28	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Phos(54) _Met(104)	14302.2	14301	70.28	10	16	26	2.24E-27	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(54) _Phos(95) _Met(104)	14302.2	14301	70.28	12	6	18	6.17E-15	plausible	1	1	1
P04370-8_Ac(1) _Phos(95) _dMet(104) _Phos(122)	14302.2	14301	70.28	14	1	15	2.75E-12	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Phos(54) _dMet(104)	14302.2	14301	70.28	10	8	18	2.24E-13	plausible	1	1	1
P04370-8_Ac(1) _Phos(54) _Phos(95) _dMet(104)	14302.2	14301	70.28	12	5	17	4.57E-12	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(47) _Phos(95) _Met(104)	14302.2	14301	70.28	10	6	16	7.08E-13	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Phos(95)	14302.2	14301	70.28	13	4	17	5.13E-12	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(54) _Met(104) _Phos(112)	14302.2	14301	70.28	12	2	14	1.05E-11	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Met(104) _Phos(112)	14302.2	14301	70.28	12	2	14	9.77E-11	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(95) _Met(104) _Phos(112)	14302.2	14301	70.28	12	2	14	1.15E-10	plausible	1	1	1
P04370-8_Ac(1) _dMet(104) _Phos(112) _Phos(122)	14302.2	14301	70.28	14	1	15	6.17E-11	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _dMet(104) _Phos(112)	14302.2	14301	70.28	10	3	13	6.31E-10	plausible	1	1	1
P04370-8_Ac(1) _Phos(54) _dMet(104) _Phos(112)	14302.2	14301	70.28	12	3	15	2.57E-09	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Phos(95) _dMet(104)	14302.2	14301	70.28	10	5	15	1.45E-09	plausible	1	1	1
P04370-8_Ac(1) _Phos(95) _dMet(104) _Phos(112)	14302.2	14301	70.28	14	2	16	1.12E-09	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Phos(95) _Met(104)	14302.2	14301	70.28	10	6	16	8.32E-10	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(54) _Phos(95) _Met(104)	14302.2	14301	70.28	12	6	18	4.27E-11	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(95) _Phos(122)	14302.2	14301	70.28	12	1	13	7.24E-09	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(104) _Phos(112) _Phos(122)	14302.2	14301	70.28	12	1	13	3.72E-08	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(54) _Met(104) _Phos(112)	14302.2	14301	70.28	12	2	14	4.27E-08	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(47) _Met(104) _Phos(112)	14302.2	14301	70.28	10	2	12	8.13E-08	plausible	1	1	1
P04370-8_Ac(1) _Phos(54) _dMet(104) _Phos(122)	14302.2	14301	70.28	12	2	14	2.57E-07	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Met(104) _Phos(112) _Phos(122)	14302.2	14301	70.28	12	1	13	1.86E-07	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _Phos(54)	14302.2	14301	70.28	10	5	15	5.5E-09	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Phos(112)	14302.2	14301	70.28	13	2	15	3.72E-07	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(54) _Met(104) _Phos(122)	14302.2	14301	70.28	12	1	13	1E-07	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Phos(122)	14302.2	14301	70.28	13	1	14	5.89E-06	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(95) _Phos(112)	14302.2	14301	70.28	12	2	14	7.59E-07	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(104) _Phos(122)	14302.2	14301	70.28	10	1	11	1.35E-06	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(95) _Met(104) _Phos(122)	14302.2	14301	70.28	12	1	13	1.38E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(47) _Met(104) _Phos(122)	14302.2	14301	70.28	10	1	11	5.37E-05	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(112) _Phos(122)	14302.2	14301	70.28	12	1	13	0.000794	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Met(104) _Phos(122)	14302.2	14301	70.28	12	1	13	8.71E-05	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(104) _Phos(112)	14302.2	14301	70.28	10	2	12	0.000141	plausible	1	1	1

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14287.1 Average									
P04370-8_Ac(1) _Phos(54) _Phos(95) _Met(104)	14287.2	14287	2.76	12	13	25	1.35E-07	plausible	1	1	1
P04370-8_Ac(1) _Phos(95) _Met(104) _Phos(122)	14287.2	14287	2.76	18	1	19	5.13E-05	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(95) _Phos(122)	14287.2	14287	2.76	13	3	16	0.000562	**valid**	2	0	2
P04370-8_Ac(1) _Phos(95) _Met(104) _Phos(112)	14287.2	14287	2.76	18	3	21	2.75E-05	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Phos(112)	14287.2	14287	2.76	13	5	18	0.000263	plausible	1	0	1
		14273.1 Average									
P04370-8_Ac(1) _Phos(95) _Phos(122)	14273.1	14273	-0.34	18	3	21	1.1E-16	**valid**	3	3	4
P04370-8_Ac(1) _Phos(112) _Phos(122)	14273.1	14273	-0.34	18	3	21	1.55E-14	**valid**	3	3	4
P04370-8_Ac(1) _Phos(95) _Phos(112)	14273.1	14273	-0.34	18	5	23	5.62E-15	**valid**	3	3	4
		14271.2 Average									
P04370-8_Ac(1) _dOx(19) _Met(104) _Phos(112) _dOx(124)	14271.1	14271	-0.55	7	1	8	0.000282	plausible	1	0	1
		14269.2 Average									
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(47) _Met(104) _Ox(124)	14269.1	14269	-2.51	7	12	19	7.08E-15	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(41) _Met(104) _Ox(124)	14269.1	14269	-2.51	7	12	19	1.12E-13	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(54) _Met(104) _Ox(124)	14269.1	14269	-2.51	8	12	20	4.07E-13	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _dMet(104) _Phos(112) _dOx(124)	14269.1	14269	-2.51	14	1	15	3.31E-09	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41) _Met(104) _Ox(124)	14269.1	14269	-2.51	8	12	20	8.32E-12	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Met(47) _dOx(124)	14269.1	14269	-2.51	11	1	12	4.79E-07	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(47) _Met(104) _Ox(124)	14269.1	14269	-2.51	8	12	20	4.07E-11	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _Phos(54) _dOx(124)	14269.1	14269	-2.51	11	1	12	0.000224	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(47) _Phos(54) _Met(104) _Ox(124)	14269.1	14269	-2.51	8	12	20	1.48E-05	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(47) _Met(104) _Ox(124)	14269.1	14269	-2.51	2	12	14	1.48E-05	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(41) _Met(104) _Ox(124)	14269.1	14269	-2.51	2	12	14	6.46E-05	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Met(104) _dOx(124)	14269.1	14269	-2.51	9	2	11	0.000871	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(95) _dMet(104) _dOx(124)	14269.1	14269	-2.51	14	1	15	0.001778	plausible	1	0	1
		14267.2 Average									
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Met(47) _Met(104) _Ox(124)	14267.2	14267	-0.38	9	12	21	3.63E-17	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _Phos(54) _Met(104) _Ox(124)	14267.2	14267	-0.38	10	12	22	3.47E-16	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Met(47) _Met(104) _Ox(124)	14267.2	14267	-0.38	7	12	19	3.89E-13	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _dMet(104) _Phos(112) _Ox(124)	14267.2	14267	-0.38	11	4	15	0.002512	**valid**	2	1	3
P04370-8_Ac(1) _Met(47) _Phos(95) _dMet(104) _dOx(124)	14267.2	14267	-0.38	13	1	14	1.12E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(41) _Met(47) _Met(104)	14267.2	14267	-0.38	7	23	30	2.24E-10	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41) _Met(47) _Met(104)	14267.2	14267	-0.38	8	23	31	9.12E-11	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _dMet(104) _Phos(122) _dOx(124)	14267.2	14267	-0.38	13	1	14	1.2E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _Met(104) _dOx(124)	14267.2	14267	-0.38	11	2	13	5.5E-08	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(47) _dMet(104) _Phos(122)	14267.2	14267	-0.38	8	2	10	4.07E-06	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Met(47) _Phos(54) _Met(104)	14267.2	14267	-0.38	8	23	31	9.12E-11	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(47) _dMet(104)	14267.2	14267	-0.38	8	3	11	5.75E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(95) _dMet(104)	14267.2	14267	-0.38	8	4	12	0.000813	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _dMet(104) _Phos(112)	14267.2	14267	-0.38	8	3	11	9.12E-06	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _dMet(104) _Phos(122)	14267.2	14267	-0.38	8	2	10	2.34E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41) _dMet(104)	14267.2	14267	-0.38	8	3	11	2.51E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(54) _dMet(104)	14267.2	14267	-0.38	8	3	11	0.001148	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _dMet(104) _Ox(124)	14267.2	14267	-0.38	9	3	12	0.000562	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Met(47) _Phos(95) _Met(104)	14267.2	14267	-0.38	8	10	18	0.001023	plausible	1	1	1
		14257.1 Average									
P04370-8_Ac(1) _dOx(19) _Phos(112) _dOx(124)	14257.2	14257	3.05	9	1	10	0.000912	plausible	1	0	1
		14255.2 Average									
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(47) _Ox(124)	14254.2	14255	-64.52	8	12	20	5.13E-12	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41) _Ox(124)	14254.2	14255	-64.52	8	13	21	3.63E-10	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(47) _Ox(124)	14254.2	14255	-64.52	7	13	20	6.92E-12	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(41) _Ox(124)	14254.2	14255	-64.52	7	14	21	8.51E-12	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(47) _Phos(54) _Ox(124)	14254.2	14255	-64.52	8	12	20	1.58E-07	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _dOx(124)	14254.2	14255	-64.52	9	2	11	1.26E-06	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(95) _Met(104) _dOx(124)	14254.2	14255	-64.52	14	1	15	3.98E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(54) _Ox(124)	14254.2	14255	-64.52	8	12	20	6.03E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(41) _Ox(124)	14254.2	14255	-64.52	2	14	16	5.13E-08	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(47) _Ox(124)	14254.2	14255	-64.52	2	13	15	0.000234	plausible	1	0	1

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14253.2 Average									
P04370-8_Ac(1) _dOx(19) _Met(47) _Phos(54) _Met(104)	14253.1	14253	-3.26	8	19	27	5.89E-16	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41) _Met(104)	14253.1	14253	-3.26	8	19	27	1.1E-16	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(54) _Met(104)	14253.1	14253	-3.26	8	19	27	4.57E-15	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Met(47) _Ox(124)	14253.1	14253	-3.26	10	12	22	4.57E-10	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(41) _Met(104)	14253.1	14253	-3.26	7	19	26	3.72E-12	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(47) _Met(104)	14253.1	14253	-3.26	7	19	26	2.4E-12	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(47) _Met(104)	14253.1	14253	-3.26	8	19	27	4.57E-11	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _dMet(104) _Phos(122) _Ox(124)	14253.1	14253	-3.26	14	2	16	2.95E-09	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(122) _dOx(124)	14253.1	14253	-3.26	11	1	12	4.47E-07	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Met(47) _Ox(124)	14253.1	14253	-3.26	8	12	20	4.79E-09	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _dMet(104) _Phos(112) _Ox(124)	14253.1	14253	-3.26	16	3	19	2.14E-05	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(47) _Met(104)	14253.1	14253	-3.26	1	19	20	6.31E-06	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _dOx(124)	14253.1	14253	-3.26	11	1	12	5.01E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _dOx(124)	14253.1	14253	-3.26	10	2	12	0.000234	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(41) _Met(47) _Ox(124)	14253.1	14253	-3.26	3	12	15	0.000457	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dOx(19) _Met(41) _Met(104)	14253.1	14253	-3.26	2	22	24	0.002188	plausible	1	0	1
		14251.2 Average									
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Met(47) _Met(104)	14251.1	14251	-5.52	9	19	28	9.55E-16	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Met(47) _Met(104)	14251.1	14251	-5.52	7	19	26	2.24E-15	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _Phos(54) _Met(104)	14251.1	14251	-5.52	9	19	28	2.51E-12	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(41) _Met(47) _Met(104)	14251.1	14251	-5.52	1	19	20	4.07E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _Met(104) _Ox(124)	14251.1	14251	-5.52	10	13	23	1.1E-11	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Met(104) _Ox(124)	14251.1	14251	-5.52	11	13	24	1.15E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Met(41) _Met(47) _Met(104) _Ox(124)	14251.1	14251	-5.52	7	13	20	3.55E-07	plausible	1	0	1
		14249.2 Average									
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _dMet(104)	14249.2	14249	-2.74	14	18	32	8.91E-24	**valid**	2	0	2
P04370-8_Ac(1) _Phos(17) _Met(41) _Met(47) _dMet(104)	14249.2	14249	-2.74	9	19	28	5.37E-17	**valid**	2	0	2
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _dMet(104)	14249.2	14249	-2.74	12	18	30	1.82E-14	**valid**	2	0	2
P04370-8_Ac(1) _Phos(7) _Met(41) _Met(47) _dMet(104)	14249.2	14249	-2.74	3	19	22	1.51E-05	**valid**	3	0	2
		14241.1 Average									
P04370-8_Ac(1) _Ox(19) _Phos(54) _dOx(124)	14241.1	14241	-1.08	11	1	12	2.19E-06	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _dOx(124)	14241.1	14241	-1.08	9	2	11	0.000891	plausible	1	0	1
		14239.2 Average									
P04370-8_Ac(1) _Ox(19) _Phos(54) _Met(104) _Ox(124)	14239.2	14239	3.1	12	12	24	8.13E-12	**valid**	2	0	2
P04370-8_Ac(1) _Phos(33) _Met(41) _dOx(124)	14239.2	14239	3.1	11	2	13	0.000138	**valid**	2	0	2
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(47)	14239.2	14239	3.1	8	9	17	4.79E-08	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(104) _Ox(124)	14239.2	14239	3.1	8	13	21	3.24E-09	**valid**	2	0	2
P04370-8_Ac(1) _Met(41) _Phos(122) _dOx(124)	14239.2	14239	3.1	12	1	13	5.01E-06	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(104) _Ox(124)	14239.2	14239	3.1	10	12	22	2.82E-08	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41) _Ox(124)	14239.2	14239	3.1	9	13	22	3.55E-08	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Phos(95) _Met(104)	14239.2	14239	3.1	10	8	18	5.01E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(47)	14239.2	14239	3.1	7	9	16	4.07E-07	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _Met(47) _Phos(54)	14239.2	14239	3.1	10	9	19	3.98E-05	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _Phos(33) _Met(41)	14239.2	14239	3.1	8	10	18	9.12E-05	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(112) _Ox(124)	14239.2	14239	3.1	12	4	16	7.94E-07	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(54)	14239.2	14239	3.1	8	9	17	3.31E-05	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(112) _Ox(124)	14239.2	14239	3.1	11	4	15	5.62E-05	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _dOx(19) _Met(41)	14239.2	14239	3.1	9	10	19	1.02E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(104) _Ox(124)	14239.2	14239	3.1	3	13	16	0.002399	**valid**	2	0	2
P04370-8_Ac(1) _dOx(19) _Met(41) _Phos(112)	14239.2	14239	3.1	8	2	10	5.89E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(41) _Ox(124)	14239.2	14239	3.1	2	14	16	0.000813	plausible	1	0	1

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14237.2 Average									
P04370-8_Ac(1) _Ox(19) _Phos(33) _dMet(104)	14237.1	14237	-2.42	10	20	30	1.05E-05	**valid**	2	0	2
P04370-8_Ac(1) _dMet(104) _Phos(112) _Ox(124)	14237.1	14237	-2.42	14	3	17	3.09E-08	plausible	1	1	1
P04370-8_Ac(1) _dMet(104) _Phos(122) _Ox(124)	14237.1	14237	-2.42	22	2	24	5.13E-08	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(54) _Met(104) _Ox(124)	14237.1	14237	-2.42	16	12	28	1.78E-10	plausible	1	0	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Met(104) _Ox(124)	14237.1	14237	-2.42	13	7	20	6.31E-08	plausible	1	0	1
P04370-8_Ac(1) _Phos(95) _dMet(104) _Ox(124)	14237.1	14237	-2.42	14	3	17	8.13E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _Ox(124)	14237.1	14237	-2.42	11	4	15	2.4E-07	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(95) _dMet(104)	14237.1	14237	-2.42	16	7	23	4.37E-07	**valid**	3	0	2
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41) _Met(47)	14237.1	14237	-2.42	8	10	18	2.45E-07	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Ox(124)	14237.1	14237	-2.42	14	6	20	1.78E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _dMet(104)	14237.1	14237	-2.42	9	20	29	5.5E-05	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(95) _Met(104)	14237.1	14237	-2.42	11	9	20	7.41E-07	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _dMet(104) _Phos(122)	14237.1	14237	-2.42	16	3	19	6.76E-07	**valid**	3	0	2
P04370-8_Ac(1) _Met(41) _Phos(95) _Met(104) _Ox(124)	14237.1	14237	-2.42	14	7	21	1.7E-06	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(104) _Phos(122) _Ox(124)	14237.1	14237	-2.42	14	2	16	0.000245	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _dMet(104)	14237.1	14237	-2.42	3	20	23	0.000457	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _Phos(54)	14237.1	14237	-2.42	11	10	21	0.002951	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Met(104) _Phos(122)	14237.1	14237	-2.42	12	1	13	0.002239	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Met(41) _Met(47) _Ox(124)	14237.1	14237	-2.42	8	7	15	0.000871	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(104) _Phos(122)	14237.1	14237	-2.42	11	1	12	0.002692	plausible	1	0	1
		14235.2 Average									
P04370-8_Ac(1) _Met(41) _Met(47) _Phos(54) _Met(104)	14235.2	14235	-0.96	14	20	34	1.86E-26	**valid**	2	0	2
P04370-8_Ac(1) _Phos(33) _Met(41) _Met(47) _Met(104)	14235.2	14235	-0.96	12	20	32	9.12E-26	**valid**	2	0	2
P04370-8_Ac(1) _Phos(17) _Met(41) _Met(47) _Met(104)	14235.2	14235	-0.96	9	21	30	2.04E-26	**valid**	2	0	2
P04370-8_Ac(1) _Phos(7) _Met(41) _Met(47) _Met(104)	14235.2	14235	-0.96	3	21	24	9.77E-20	**valid**	3	0	2
P04370-8_Ac(1) _Met(47) _Phos(95) _dMet(104)	14235.2	14235	-0.96	13	8	21	1.91E-16	plausible	1	0	1
P04370-8_Ac(1) _Met(47) _dMet(104) _Phos(122)	14235.2	14235	-0.96	13	1	14	4.79E-07	plausible	1	0	1
P04370-8_Ac(1) _Met(47) _dMet(104) _Phos(112)	14235.2	14235	-0.96	13	3	16	1.82E-10	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(95) _dMet(104)	14235.2	14235	-0.96	13	8	21	4.57E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Met(41) _dMet(104)	14235.2	14235	-0.96	1	19	20	2.09E-15	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Met(47) _dMet(104)	14235.2	14235	-0.96	1	19	20	6.76E-12	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _dMet(104) _Phos(112)	14235.2	14235	-0.96	13	3	16	5.13E-05	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _dMet(104) _Phos(122)	14235.2	14235	-0.96	13	1	14	0.000148	plausible	1	0	1
		14225.1 Average									
P04370-8_Ac(1) _Ox(19) _Phos(54) _Ox(124)	14224.2	14225	-64.86	12	12	24	4.37E-13	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Phos(33) _Ox(124)	14224.2	14225	-64.86	10	13	23	2.51E-09	**valid**	2	0	2
P04370-8_Ac(1) _Phos(17) _Ox(19) _Ox(124)	14224.2	14225	-64.86	8	10	18	2.34E-07	**valid**	2	0	2
P04370-8_Ac(1) _Phos(122) _dOx(124)	14224.2	14225	-64.86	15	1	16	2.14E-09	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(122) _Ox(124)	14224.2	14225	-64.86	14	1	15	6.61E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Ox(124)	14224.2	14225	-64.86	3	13	16	5.75E-06	**valid**	2	0	2
		14223.2 Average									
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(104)	14223.1	14223	-3	12	22	34	1.1E-18	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(104)	14223.1	14223	-3	9	19	28	1.2E-18	plausible	1	0	1
P04370-8_Ac(1) _Phos(95) _Met(104) _Ox(124)	14223.1	14223	-3	20	7	27	1.95E-14	plausible	1	1	1
P04370-8_Ac(1) _Phos(54) _Met(104) _Ox(124)	14223.1	14223	-3	13	14	27	4.47E-15	plausible	1	0	1
P04370-8_Ac(1) _Phos(33) _Met(41) _Ox(124)	14223.1	14223	-3	11	13	24	3.8E-11	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(47)	14223.1	14223	-3	11	9	20	8.32E-12	plausible	1	1	1
P04370-8_Ac(1) _Met(104) _Phos(112) _Ox(124)	14223.1	14223	-3	18	1	19	1.15E-11	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(122) _Ox(124)	14223.1	14223	-3	13	2	15	1.51E-10	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(54) _Met(104)	14223.1	14223	-3	13	22	35	3.09E-13	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(33) _Met(41)	14223.1	14223	-3	8	19	27	1.12E-09	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(95) _Met(104)	14223.1	14223	-3	13	8	21	2.4E-12	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(122) _Ox(124)	14223.1	14223	-3	14	2	16	3.02E-08	**valid**	2	0	2
P04370-8_Ac(1) _Phos(17) _Met(104) _Ox(124)	14223.1	14223	-3	7	14	21	5.75E-11	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(54)	14223.1	14223	-3	8	19	27	3.16E-09	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(104) _Phos(112)	14223.1	14223	-3	13	3	16	2.57E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(47)	14223.1	14223	-3	7	19	26	4.9E-09	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(54)	14223.1	14223	-3	13	9	22	1.66E-08	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Ox(124)	14223.1	14223	-3	13	4	17	5.75E-08	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(95) _Ox(124)	14223.1	14223	-3	14	4	18	7.08E-09	plausible	1	0	1
P04370-8_Ac(1) _Met(47) _Phos(54) _Ox(124)	14223.1	14223	-3	14	2	16	4.57E-08	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(95)	14223.1	14223	-3	12	8	20	4.79E-11	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(104) _Phos(122)	14223.1	14223	-3	16	1	17	6.46E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(104)	14223.1	14223	-3	1	19	20	5.25E-08	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(112)	14223.1	14223	-3	12	4	16	1.2E-09	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Ox(19) _Met(41)	14223.1	14223	-3	9	10	19	5.89E-08	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(47) _Ox(124)	14223.1	14223	-3	10	12	22	3.31E-08	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(95)	14223.1	14223	-3	12	8	20	8.51E-11	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Met(41) _Ox(124)	14223.1	14223	-3	8	14	22	7.76E-06	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(122)	14223.1	14223	-3	12	2	14	1.82E-07	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Phos(112)	14223.1	14223	-3	12	4	16	1.78E-07	plausible	1	0	1
P04370-8_Ac(1) _Phos(17) _Met(47) _Ox(124)	14223.1	14223	-3	7	13	20	2.45E-09	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _Phos(122)	14223.1	14223	-3	13	2	15	1.58E-05	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(47)	14223.1	14223	-3	2	19	21	2.19E-06	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Ox(19) _Met(41)	14223.1	14223	-3	2	19	21	0.001585	plausible	1	1	1

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14221.2 Average									
P04370-8_Ac(1) _Phos(95) _dMet(104)	14221.2	14221	1.69	18	8	26	3.16E-20	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(95) _Met(104)	14221.2	14221	1.69	13	12	25	5.75E-18	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(104) _Phos(112)	14221.2	14221	1.69	11	4	15	4.27E-07	**valid**	2	0	2
P04370-8_Ac(1) _dMet(104) _Phos(122)	14221.2	14221	1.69	18	1	19	2.95E-12	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(95) _Met(104)	14221.2	14221	1.69	11	11	22	0.000132	**valid**	2	0	2
P04370-8_Ac(1) _dMet(104) _Phos(112)	14221.2	14221	1.69	18	3	21	1.91E-12	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Met(104) _Phos(122)	14221.2	14221	1.69	11	2	13	1.15E-08	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(104) _Phos(122)	14221.2	14221	1.69	11	4	15	0.001288	**valid**	2	0	2
P04370-8_Ac(1) _Met(47) _Met(104) _Phos(112)	14221.2	14221	1.69	11	5	16	0.000347	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _dMet(104)	14221.2	14221	1.69	1	18	19	1.26E-09	plausible	1	0	1
		14209.2 Average									
P04370-8_Ac(1) _Phos(112) _Ox(124)	14209.2	14209	0.66	20	2	22	9.12E-15	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(112)	14209.2	14209	0.66	13	4	17	6.92E-11	**valid**	2	0	2
P04370-8_Ac(1) _Phos(122) _Ox(124)	14209.2	14209	0.66	20	2	22	2.34E-08	**valid**	2	2	4
P04370-8_Ac(1) _Ox(19) _Phos(54)	14209.2	14209	0.66	8	19	27	4.57E-19	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Phos(95)	14209.2	14209	0.66	13	8	21	5.62E-14	**valid**	2	0	2
P04370-8_Ac(1) _Ox(19) _Phos(33)	14209.2	14209	0.66	8	19	27	7.08E-20	plausible	1	1	1
P04370-8_Ac(1) _Phos(95) _Ox(124)	14209.2	14209	0.66	20	4	24	3.55E-15	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Ox(19)	14209.2	14209	0.66	7	19	26	1.62E-18	plausible	1	1	1
P04370-8_Ac(1) _Phos(54) _Ox(124)	14209.2	14209	0.66	13	6	19	2E-13	plausible	1	0	1
P04370-8_Ac(1) _Phos(33) _Ox(124)	14209.2	14209	0.66	11	7	18	3.63E-10	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Phos(122)	14209.2	14209	0.66	13	2	15	1.45E-05	**valid**	2	0	2
P04370-8_Ac(1) _Phos(17) _Ox(124)	14209.2	14209	0.66	7	7	14	7.41E-06	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Ox(19)	14209.2	14209	0.66	2	19	21	2.82E-09	plausible	1	1	1
		14207.2 Average									
P04370-8_Ac(1) _Phos(95) _Met(104)	14207.2	14207	-0.06	18	12	30	1.74E-23	**valid**	3	3	4
P04370-8_Ac(1) _Met(104) _Phos(112)	14207.2	14207	-0.06	18	5	23	6.46E-12	**valid**	3	3	4
P04370-8_Ac(1) _Met(104) _Phos(122)	14207.2	14207	-0.06	21	4	25	3.47E-10	**valid**	4	4	4
P04370-8_Ac(1) _Phos(7) _Met(104)	14207.2	14207	-0.06	1	20	21	1.91E-18	plausible	1	0	1
P04370-8_Ac(1) _Met(47) _Phos(95)	14207.2	14207	-0.06	11	9	20	3.98E-12	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(95)	14207.2	14207	-0.06	13	6	19	1.55E-09	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(54)	14207.2	14207	-0.06	13	21	34	4.57E-16	plausible	1	1	1
P04370-8_Ac(1) _Met(41) _Phos(112)	14207.2	14207	-0.06	11	4	15	1.48E-07	**valid**	2	0	2
P04370-8_Ac(1) _Met(47) _Phos(54)	14207.2	14207	-0.06	13	21	34	5.37E-12	plausible	1	1	1
P04370-8_Ac(1) _Phos(33) _Met(41)	14207.2	14207	-0.06	11	21	32	3.09E-14	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Met(41)	14207.2	14207	-0.06	9	21	30	1.91E-12	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(112)	14207.2	14207	-0.06	11	4	15	1.07E-08	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Phos(122)	14207.2	14207	-0.06	11	3	14	9.12E-06	**valid**	2	0	2
P04370-8_Ac(1) _Phos(33) _Met(47)	14207.2	14207	-0.06	11	21	32	2.4E-13	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _Phos(122)	14207.2	14207	-0.06	13	1	14	3.8E-08	plausible	1	0	1
P04370-8_Ac(1) _Phos(7) _Met(41)	14207.2	14207	-0.06	3	21	24	9.12E-10	plausible	1	1	1
P04370-8_Ac(1) _Phos(7) _Met(47)	14207.2	14207	-0.06	3	21	24	6.03E-09	plausible	1	1	1
P04370-8_Ac(1) _Phos(17) _Met(47)	14207.2	14207	-0.06	9	9	18	4.37E-05	plausible	1	1	1
		14205.2 Average									
P04370-8_Ac(1) _dOx(19) _Met(41) _Met(104) _dOx(124)	14205.2	14205	-0.14	9	2	11	1.32E-06	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(47) _Met(104) _dOx(124)	14205.2	14205	-0.14	9	2	11	1.82E-05	plausible	1	0	1
		14201.2 Average									
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _dMet(104) _Ox(124)	14201.1	14201	-7.88	7	0	7	0.000437	plausible	1	1	1
		14193.2 Average									
P04370-8_Ac(1) _Phos(95)	14192.2	14193	-64.5	17	9	26	1.41E-20	**valid**	3	3	4
P04370-8_Ac(1) _Phos(122)	14192.2	14193	-64.5	21	3	24	3.16E-16	**valid**	3	3	4
P04370-8_Ac(1) _Phos(112)	14192.2	14193	-64.5	21	4	25	1.51E-17	**valid**	3	3	4
P04370-8_Ac(1) _Phos(17)	14192.2	14193	-64.5	7	19	26	2.29E-18	plausible	1	1	1
P04370-8_Ac(1) _Phos(7)	14192.2	14193	-64.5	2	21	23	2.95E-17	plausible	1	1	1
		14191.2 Average									
P04370-8_Ac(1) _dOx(19) _Met(41) _dOx(124)	14191.2	14191	0.37	9	1	10	3.09E-05	plausible	1	0	1
		14189.2 Average									
P04370-8_Ac(1) _dOx(19) _dMet(104) _Ox(124)	14189.2	14189	0.04	10	2	12	6.61E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Met(104) _Ox(124)	14189.2	14189	0.04	8	6	14	3.39E-05	plausible	1	0	1
		14187.2 Average									
P04370-8_Ac(1) _Met(41) _dMet(104) _dOx(124)	14187.1	14187	-6.02	14	1	15	6.17E-09	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _dMet(104) _Ox(124)	14187.1	14187	-6.02	12	2	14	2.14E-08	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _dMet(104) _Ox(124)	14187.1	14187	-6.02	12	2	14	4.57E-07	plausible	1	0	1
		14175.2 Average									
P04370-8_Ac(1) _Ox(19) _Met(41) _dOx(124)	14175.2	14175	1.64	13	2	15	1.66E-05	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(47) _dOx(124)	14175.2	14175	1.64	12	1	13	1.2E-05	plausible	1	0	1
P04370-8_Ac(1) _dOx(19) _Met(41) _Ox(124)	14175.2	14175	1.64	8	13	21	2.29E-05	plausible	1	0	1
		14173.2 Average									
P04370-8_Ac(1) _Ox(19) _dMet(104) _Ox(124)	14173.2	14173	-0.05	13	2	15	7.59E-09	**valid**	2	0	2
P04370-8_Ac(1) _dMet(104) _dOx(124)	14173.2	14173	-0.05	18	1	19	3.63E-09	plausible	1	1	1
P04370-8_Ac(1) _dOx(19) _dMet(104)	14173.2	14173	-0.05	10	5	15	9.77E-08	plausible	1	0	1

Proteoform Identifier	Obs Mass	Thr Mass	Mass Error (ppm)	b ions	y ions	Frag Total	P-Score (adj)	Confirmation	Min. # Frags PTM	# Frags Confirm All Internal PTM	Rank*
		14171.2 Average									
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(47) _Met(104)	14171.2	14171	-5.3	8	19	27	2.04E-19	plausible	1	1	1
P04370-8_Ac(1) _Met(47) _dMet(104) _Ox(124)	14171.2	14171	-5.3	14	3	17	2.04E-05	**valid**	2	1	3
P04370-8_Ac(1) _Met(41) _dMet(104) _Ox(124)	14171.2	14171	-5.3	15	3	18	3.8E-07	**valid**	2	1	3
P04370-8_Ac(1) _Met(41) _Met(47) _Met(104) _Ox(124)	14171.2	14171	-5.3	14	11	25	4.37E-16	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(47) _dMet(104)	14171.2	14171	-5.3	8	3	11	5.5E-05	plausible	1	1	1
		14169.3 Average									
P04370-8_Ac(1) _Met(41) _Met(47) _dMet(104)	14170.1	14169	60.96	12	19	31	4.27E-18	plausible	1	1	1
		14161.2 Average									
P04370-8_Ac(1) _Ox(19) _dOx(124)	14161.1	14161	-2.37	14	2	16	6.61E-07	**valid**	2	0	2
		14159.2 Average									
P04370-8_Ac(1) _Ox(19) _Met(104) _Ox(124)	14159.2	14159	-0.28	14	13	27	7.76E-09	**valid**	7	0	2
P04370-8_Ac(1) _Ox(19) _Met(47) _Ox(124)	14159.2	14159	-0.28	10	13	23	1.95E-12	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Ox(124)	14159.2	14159	-0.28	9	14	23	2.4E-12	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _dOx(124)	14159.2	14159	-0.28	12	2	14	5.01E-08	**valid**	2	0	2
P04370-8_Ac(1) _Met(47) _dOx(124)	14159.2	14159	-0.28	13	1	14	3.16E-08	plausible	1	0	1
		14157.2 Average									
P04370-8_Ac(1) _dMet(104) _Ox(124)	14157.2	14157	-2.53	18	2	20	6.46E-16	**valid**	2	2	4
P04370-8_Ac(1) _Ox(19) _Met(47) _Met(104)	14157.2	14157	-2.53	13	22	35	2.63E-21	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _Met(41) _Met(104)	14157.2	14157	-2.53	13	22	35	1.07E-20	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(104) _Ox(124)	14157.2	14157	-2.53	14	14	28	1.74E-11	**valid**	3	0	2
P04370-8_Ac(1) _Met(47) _Met(104) _Ox(124)	14157.2	14157	-2.53	13	13	26	3.89E-12	plausible	1	0	1
P04370-8_Ac(1) _Ox(19) _dMet(104)	14157.2	14157	-2.53	13	5	18	2.24E-13	plausible	1	0	1
P04370-8_Ac(1) _Met(41) _Met(47) _Ox(124)	14157.2	14157	-2.53	13	7	20	2.75E-10	plausible	1	0	1
		14155.2 Average									
P04370-8_Ac(1) _Met(41) _Met(47) _Met(104)	14155.1	14155	-8.15	12	22	34	5.5E-30	**valid**	2	2	4
P04370-8_Ac(1) _Met(41) _dMet(104)	14155.1	14155	-8.15	13	20	33	4.68E-27	**valid**	4	2	4
P04370-8_Ac(1) _Met(47) _dMet(104)	14155.1	14155	-8.15	13	20	33	3.98E-29	**valid**	3	2	4
		14145.2 Average									
P04370-8_Ac(1) _Ox(19) _Ox(124)	14145.2	14145	-0.16	13	13	26	3.09E-14	**valid**	6	0	2
P04370-8_Ac(1) _dOx(124)	14145.2	14145	-0.16	15	2	17	2.88E-06	**valid**	2	2	4
		14143.2 Average									
P04370-8_Ac(1) _Met(104) _Ox(124)	14143.2	14143	1.07	20	14	34	1.38E-31	**valid**	12	12	4
P04370-8_Ac(1) _Ox(19) _Met(104)	14143.2	14143	1.07	13	19	32	9.12E-22	**valid**	4	0	2
P04370-8_Ac(1) _Met(47) _Ox(124)	14143.2	14143	1.07	13	7	20	7.59E-14	**valid**	2	1	3
P04370-8_Ac(1) _Met(41) _Ox(124)	14143.2	14143	1.07	12	14	26	3.98E-13	**valid**	3	1	3
P04370-8_Ac(1) _Ox(19) _Met(47)	14143.2	14143	1.07	13	11	24	1.32E-11	plausible	1	1	1
P04370-8_Ac(1) _Ox(19) _Met(41)	14143.2	14143	1.07	12	10	22	1.91E-10	plausible	1	1	1
		14141.2 Average									
P04370-8_Ac(1) _dMet(104)	14141.2	14141	-1.02	18	20	38	4.37E-25	**valid**	16	16	4
P04370-8_Ac(1) _Met(41) _Met(104)	14141.2	14141	-1.02	11	21	32	3.39E-31	**valid**	2	0	2
P04370-8_Ac(1) _Met(47) _Met(104)	14141.2	14141	-1.02	11	21	32	1.7E-29	plausible	1	0	1
		14129.2 Average									
P04370-8_Ac(1) _Ox(124)	14129.1	14129	-4.18	20	7	27	7.76E-22	**valid**	7	7	4
P04370-8_Ac(1) _Ox(19)	14129.1	14129	-4.18	16	11	27	4.9E-10	**valid**	8	8	4
		14127.2 Average									
P04370-8_Ac(1) _Met(104)	14127.2	14127	-0.26	21	24	45	6.17E-49	**valid**	21	21	4
P04370-8_Ac(1) _Met(47)	14127.2	14127	-0.26	11	6	17	7.76E-09	**valid**	2	2	4
P04370-8_Ac(1) _Met(41)	14127.2	14127	-0.26	12	21	33	8.32E-11	plausible	1	1	1
		14113.2 Average									
P04370-8_Ac(1)	14113.1	14113	-4.54	19	23	42	6.31E-40	**valid**	19	19	4

APPENDIX V ISOLATION OF MBP FROM RODENT TISSUE

Isolation of MBP from rodent brain and spine tissue

Items

1. Chem Wipes
2. 70% ethanol in water
3. Pipettes
4. Pipette tips
5. Ice bucket / wet ice
6. Glass vials
7. Sonicator
8. Vortex
9. 1.5 mL Protein low-bind Eppendorf microcentrifuge tubes
10. 0.6 mL Protein low-bind Eppendorf microcentrifuge tubes
11. Microcentifuge tube racks
12. Liquid nitrogen dewer / liquid nitrogen
13. Benchtop microcentrifuge

Animal Sacrifice

1. When sacrificing the animals, perfuse them with PBS prior to tissue collection via cardiac puncture to reduce the amount of blood collected in the tissue. The use of isoflurane and nitrous oxide to anesthetize and asphyxiate the animals may promote oxidation, but that was not thoroughly tested versus alternatives such as decapitation.

Tissue Storage

1. Flash freeze tissue immediately after isolation. Store the tissue at -80 C. Conduct the MBP extraction within a few days of when are ready to analyze the MBP to prevent artefactual oxidation of isolated MBP in storage. MBP oxidizes faster after it has been isolated than when it is still in the tissue.

Clean the Sonicator

1. Use chemwipes with 70% ethanol in water to clean the probe of the sonicator.

Solvent Preparation

1. Prepare the alcohol/water/acid solvent by adding 4 mL water, 4 mL isopropyl alcohol, and 250 uL of 2N HCl in a glass vial. Use optima grade (Fisher Scientific) water and IPA. Avoid using plastic pipettes to transfer the acid, use a syringe or glass pipette instead.
2. Prepare Chloroform by adding it to a glass bottle
3. Prepare Water by adding it to a glass bottle.

4. Keep all of these solvents on wet ice.

All glass bottles used must be thoroughly cleaned by sonication when full of IPA prior to use. Bottles used for media or other purposes may be too contaminated even after cycling through a glass washer. Best practice would be to use new bottles that have been cleaned via IPA sonication.

MBP Isolation Procedure

1. Use the sonicator in a cold room if possible.
2. Cut the tissue into 1-2 mm^3 squares on a cold, metal block prior to extraction using a scalpel or razor blade. This reduces the duration of sonication necessary to disrupt the tissue which serves to reduce sonication. Using a cold block prevents the tissue from unthawing, and facilitates making clean cuts. If the tissue becomes warm it becomes gel-like and is difficult to cut. Use <200 mg total tissue per 1.5 mL protein low-bind Eppendorf tube.
3. Put the tissue fragments into a 1.5 mL protein low-bind tube. Add 1 mL of 4 C chloroform.
4. Sonicate the tissue using pulses of ~1 second on ~1 second off. The tissue should disrupt in less than 30 seconds. The liquid will have a foamy, white appearance after sonication. Stop sonicating when the tissue has been disrupted if this occurs in less than 30 seconds as the sonication can facilitate oxidation. Clean the probe with 70% ethanol in water and Chemwipes after each sample.
5. If samples from different biological states are being prepared, take care to sonicate for the same time duration.
6. Utilize the same probe depth (the depth into the chloroform in which the probe is submerged) when sonicating. Sonication is more intense when the probe is less submerged.
7. Spin the sample at max speed (14,000 RCF) on a benchtop microcentrifuge for 10 minutes. Take care in removing the sample from the centrifuge, excess motion can disrupt the layers of the separation. After spinning the sample, there will be an aqueous layer on top, a layer of tissue in the middle, and chloroform at the bottom.
8. Carefully remove and discard the aqueous layer
9. Add 200 uL 4 C water to wash
10. Sonicate the tissue again as above.
11. Centrifuge the sample again as above
12. Carefully remove the tube from the microcentrifuge. Remove the (top) aqueous layer.

13. Carefully remove the layer of cellular debris using a pipet tip. It is not necessary to remove every fragment of tissue. Sometimes the layer will dissociate easily, remove as much as possible with minimal agitation of the interface between the chloroform (bottom) and tissue.
14. Remove the bottom ~70% of the chloroform layer while avoiding touching or pipetting any cellular debris or remainder of any aqueous phase. Capture as much as possible while avoiding any contamination from the cellular debris or aqueous liquid above the interface.
15. Add an equal volume of 4 C water/IPA/HCl to the chloroform
16. Vortex
17. Centrifuge the sample again as above
18. Extract the top of the aqueous layer, taking care not to pipet near the interface of the aqueous layer / tissue remainder / chloroform.
19. Aliquot the aqueous layer into 0.6 mL protein low-bind microcentrifuge tubes as desired.
20. Flash freeze the aqueous layer. Prepare the liquid nitrogen dewer before extraction to minimize the time the MBP is in the water phase prior to lyophilization.
21. Lyophilize the frozen samples.
22. Store at -80 for ~2-3 days before LC/MS analysis.

MS / HPLC tips

1. Prepare fresh mobile phase solvents at least every 7-10 days
2. Degas LC solvents by sonication, vacuum sonication, or purging via inert gas.
3. Purge the flowmeter and pump blocks thoroughly
4. Minimize the time between reconstitution of the sample and injection. Oxidation rapidly occurs and room temperature or even on wet ice, even over a matter of minutes.

APPENDIX VI ACKNOWLEDGEMENTS

I would like to thank my PI, Dr. Steven Patrie, my thesis committee, Dr. Yonghao Yu (chair), Dr. Nancy Monson, and Dr. Nick Grishin. I would like to thank former scientists within the group, Dr. Michael Roth and Dr. Junmei Zhang. I would like to thank Casey Wing for her work on the Excel Macro Workflow and MBP isolation, and John Corbett for feedback. I would also like to thank all former lab members of the Patrie Lab. I thank the Molecular Biophysics community and Computational and Systems Biology community at UT Southwestern for advice and input. For work with SPLC, the authors are grateful to Drs. Michael Story and Deepti Ramnarain for HeLa cells, Drs. Haydn Ball and Yan Li for use of the LTQ Orbitrap Velos, and Dr. James Stull for help with development of whole heart samples. We give special thanks Maureen Joseph and the many others with Agilent Technologies that enabled this work. Thanks also to Amanda Berg and Gary Valaskovic with New Objective, Inc. For computation and statistical support I would like to thank Richard LeDuc, Ryan Fellers, and Dana Robinson.