Proteoform-selective imaging of tissues using mass spectrometry

Manxi Yang[a], Hang Hu[a], Pei Su[a,b], Paul M. Thomas[b], Jeannie M. Camarillo[b], Joseph B. Greer[b], Bryan P. Early[b], Ryan T. Fellers[b], Neil L. Kelleher*[b], Julia Laskin*[a]

[a] M. Yang, H. Hu, P. Su, Dr. J. Laskin
Department of Chemistry, Purdue University
560 Oval Drive, West Lafayette, IN 47907 (USA)
E-mail: jlaskin@purdue.edu

[b] P. M. Thomas, J.M. Camarillo, B. P. Early, R. T. Fellers, J. B. Greer, P. Su, Dr. N. L. Kelleher
Departments of Chemistry and Molecular Biosciences, Northwestern University
2145 Sheridan Road, Evanston, IL, 60208 (USA)

Supporting information for this article is given via a link at the end of the document.

Abstract: *Unraveling the complexity of biological systems relies on the development of new approaches for spatially resolved proteoform-specific analysis of the proteome. Top-down proteomics is a powerful tool, which has been used for the identification of thousands of proteoforms in biological samples. Herein, we present a first spatially resolved top-down proteomics analysis of biological tissues using nanospray desorption electrospray ionization mass spectrometry imaging (nano-DESI MSI). Nano-DESI generates multiply charged protein ions, which is advantageous for their structural characterization using tandem mass spectrometry (MS/MS). Proof-of-concept experiments demonstrate that the nano-DESI MSI combined with on-tissue top-down proteomics is ideally suited for the proteoform-selective imaging of thin tissue sections. Using rat brain tissue as a model system, we provide the first evidence of the differential proteoform expression in different regions of the brain.*

A proteome is a collection of proteins expressed by an organism, which determines its biological state. Post-translational modification (PTM) of proteins is a process that dramatically increases the complexity of the proteome beyond what can be predicted based on the known protein-coding genes.[1][2] PTMs affect protein structure, interactions, transport, and function thereby exhibiting a pronounced effect on cell physiology.[3] Furthermore, the reversible nature of PTMs facilitates the dynamic response of biological systems to different conditions in both health and disease. Therefore, the identification of proteoforms is critical to understanding biological processes.

Mass-spectrometry-based top-down proteomics is a powerful tool for proteoform characterization. This approach enables the identification of intact proteoforms based on their mass-to-charge ratio (*m/z*) and fragmentation pattern.[4] Top-down proteomics has been used for the identification of thousands of proteoforms in biological samples.[5][6][7][8] Several databases have been developed as a community resource for proteoform identification.[9][10] Several studies used top-down proteomics for the discovery of disease biomarkers.[11] For example, differences in the expression levels of several KRAS proteoforms have been associated with cancer cell proliferation.[12] In another study, the phosphorylated cardiac troponin I was identified as a potential biomarker of chronic heart failure.[13]

Although top-down proteomics is an established field, mapping of the proteoform localization in tissues is still challenging. Mass spectrometry imaging (MSI) is a label-free approach for the untargeted analysis of biomolecules with little sample preparation, which is ideally suited for spatial proteomics studies. MSI may be used for determining the spatial localization of multiple proteoforms in a single experiment, making it a powerful discovery tool for biological research.

Matrix-assisted laser desorption ionization (MALDI) has been used for imaging of intact proteins smaller than 22 kDa.[14][15] Although a majority if these studies were performed with a moderate spatial resolution of 50 µm, a subcellular spatial resolution has been achieved using transmission-geometry MALDI source.[16] Despite these developments, the low charge states of ions generated in MALDI presents a challenge to the detection of proteins using commercial mass spectrometers and their identification using top-down tandem mass spectrometry (MS/MS). The latter limitation has been addressed by combining MALDI MSI with the off-line protein identification using conventional top-down approaches.[17] Alternatively, on-tissue protein digestion combined with bottom-up proteomics has been used for protein imaging.[18] Nanodroplet processing in one pot for trace samples (nanoPOTs)[19] is an emerging bottom-up technique that does not require on-tissue digestion. NanoPOTS using laser capture microdissection (LCM) enables imaging and identification of proteins in tissue sections with a spatial resolution of 50-100 µm.[20]

Ambient ionization based on liquid extraction generate higher charge states of protein ions making these techniques suitable for conducting top-down proteomics directly from a tissue sample. Desorption electrospray ionization (DESI) and liquid extraction surface analysis (LESA) have been used for detecting intact proteins in tissues.[21] DESI has been used for imaging of <20 kDa proteins with a spatial resolution of 150-200 µm.[22][23] Meanwhile, the localization of larger proteins and protein complexes of up to 47 kDa has been examined using LESA with a spatial resolution of 600 µm.[24][25]

Nanospray desorption electrospray ionization (nano-DESI) is an ambient ionization technique, which has been extensively used for imaging of lipids and metabolites with high sensitivity and a spatial resolution down to 10 µm.[26] Although the utility of nano-DESI for protein analysis has been demonstrated[27], only a few studies used it for protein imaging. In particular, nano-DESI MSI has been used for protein localization in healthy and MYC-induced lymphomas mouse brain tissue sections[28], imaging of proteins in skin melanoma,[29] and native MSI of protein complexes in rat kidney tissues.[30]

Despite the impressive developments in spatial proteomics, the direct imaging and identification of proteoforms has not advanced significantly in recent years. In this study, we utilize nano-DESI MSI combined with top-down proteomics directly on a tissue for imaging of proteoforms. Proof-of-concept experiments presented in this study used rat brain tissue sections to demonstrate the ability of nano-DESI MSI to generate high-quality proteoform-specific imaging data. Each proteoform with unique PTMs was selected to generate the ion image, thus we name our workflow as proteoform-selective MSI. We observed forty proteoforms of nineteen <19 kDa proteins and investigated the relative abundance of multiple proteoforms of the 14.1 kDa myelin basic protein (MBP). Our results provide first insights into the differences in proteoform expression levels in different regions of the brain, which is important for understanding the relevant biological pathways.
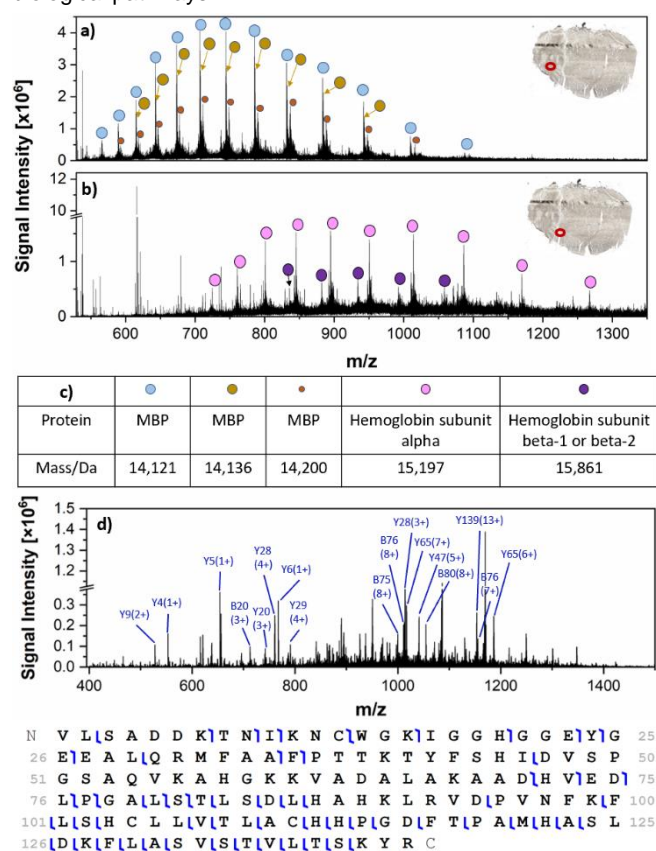

Figure 1. Average mass spectra of the rat brain tissue showing protein signals in the (a) white mater and (b) blood vessel regions highlighted with red circles in the corresponding optical images. (c) A table summarizing the features labeled in a) and b). (d) MS/MS spectrum of the unmodified hemoglobin subunit alpha (HBA1) and its fragmentation map.

Nano-DESI MSI experiments and on-tissue top-down proteomics are described in detail in the supporting information (SI). Briefly, rat brain tissue sections were delipidated by sequential washes in the ethanol solutions (70%, 90%, and 100%) and chloroform right before the analysis. Proteoforms were extracted into the liquid bridge formed between two fused capillaries and ionized by electrospray ionization at a mass spectrometer inlet. ACN/H_2O/CH_3COOH (65/34/1, v/v/v) was used as the extraction solvent. On-tissue top-down proteomics analysis was performed using targeted MS/MS. Proteoforms were identified by matching their intact masses and MS/MS data against the Rattus norvegicus (Rat: Taxon 10116) database.

Representative mass spectra from two regions of the brain tissue are shown in Figure 1. The 14.1 kDa MBP is the most abundant species detected in the white matter region of the cerebellum (Figure 1a). We observe the unmodified 14.1 kDa MBP and 10 PTMs with the charge state distribution in the range of +15 to +26. The most abundant proteoforms and their intact masses are listed in Figure 1c. We also detected an 18.4 kDa isoform of MBP and its two PTMs in the white matter region. In addition to MBPs, we observe abundant signals of hemoglobin proteoforms in the blood vessel region (Figure 1b). The mass spectrum obtained for this region contains the charge state distribution of hemoglobin subunit alpha (HBA1), subunit beta-1 (HBB1), and beta-2 (HBB2) with a total of 13 proteoforms in the range of +11 to +20. The top two most abundant proteoforms are listed in Figure 1c. Figure 1d shows the annotated MS/MS spectrum of the unmodified HBA1 and its corresponding fragmentation map with 41% sequence coverage. Most of the other proteoforms detected in this study are enhanced in the hippocampal formation region. A representative averaged mass spectrum obtained from this region is shown in Figure S2. A complete list of the proteoforms identified and imaged in this study with their proteoform identification levels is provided in the supporting table.[31]

Ion images of several abundant proteoforms normalized to the total ion current (TIC) are shown in Figure 2. Additional images are provided in Figure S3. Ion images of the same proteoforms obtained in replicate experiments are shown in Figures S4 and S5. The two most abundant proteoforms of hemoglobin, unmodified HBA1 (Figure 2b) and HBB-1 or HBB-2 (Figure 2c), are enhanced in the blood vessel region. Note that there is not sufficient fragmentation to accurately assign Figure 2c to HBB-1 or HBB-2. The fragment ions captured in MS/MS experiments matched well with the overlapped C-terminal amino
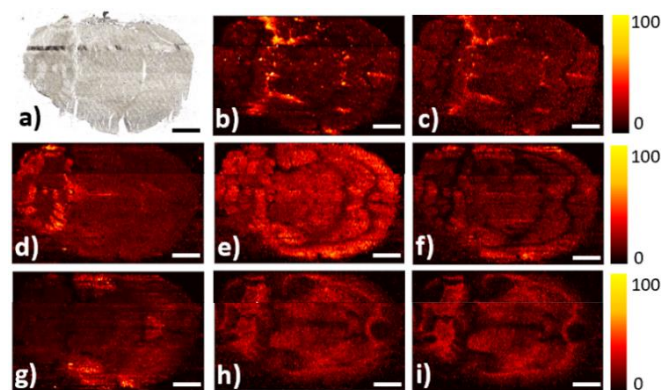

Figure 2. (a) An optical image of an axial rat brain tissue section after nano-DESI MSI analysis. Ion images of the intact proteins normalized to TIC: (b) *m/z* 894.9384[17+], 15,197Da, Hemoglobin subunit alpha (HBA1), unmodified, (c) *m/z* 992.281[16+], 15,861 Da, Hemoglobin subunit beta-1 (HBB-1) or Hemoglobin subunit beta-2 (HBB-2), (d) *m/z* 765.4067[13+], 9,937 Da, Acyl-CoA-binding protein (ACBP), N-terminal or K7 acetylation, (e) *m/z* 784.564[7+], 5,484 Da, Cytochrome c oxidase subunit 7c, mitochondrial (COX7c), unmodified, (f) *m/z* 998.7947[10+], 9,978 Da, Cytochrome c oxidase subunit 6B1 (COX6B1), N-terminal acetylation, (g) *m/z* 709.9375[7+], 4,962 Da, Thymosin beta-4, N-terminal, K3 or K11 acetylation, (h) *m/z* 741.8094[19+], 14,075 Da, myelin basic protein (MBP), unmodified, (i) *m/z* 682.3572[27+], 18,397 Da, myelin basic protein (MBP), N-terminal acetylation. Scale bar: 3mm.

acids sequence of HBB-1 and HBB-2. Similar localization is observed for all the hemoglobin proteoforms. The observed localization of hemoglobin proteoforms to the blood vessel region validates the proteoform-selective imaging experiments. Additional ion images of hemoglobin in replicate experiments are shown in Figures S10, S11, and S12.

Figure 2d shows the unique distribution of the acetylated acyl-CoA binding protein (ACBP) enhanced in the grey matter region of the rat brain cerebellum. ACBP functions as a transporter of both medium- and long-chain acyl-CoA esters. The observed enhancement of the ACBP in the grey matter is consistent with its previously reported localization to the glial cells and synaptosomal nerve ending, the major components of the grey matter in rat brain.[32] Both the unmodified mitochondrial cytochrome c oxidase, subunit 7c (Figure 2e) and N-terminal acetylated cytochrome c oxidase 6B1 (Figure 2f) are evenly distributed across the rat brain tissue. Figure 2g displays the localization of the acetylated 4.9 kDa thymosin beta-4, which is enhanced in the hippocampal formation, isocortex, lateral spetal nucles, and striatum ventral regions. This distribution is consistent with the high expression level of thymosin beta-4 messenger RNA (mRNA) in the hippocampal formation and cerebral cortex regions.[33][34] A similar localization is observed for the acetylated 4.7 kDa isoform of thymosin beta-4 (Figure S4t).

Ion images of the two isoforms of MBP, unmodified 14.1 kDa MBP and N-terminal acetylated 18.4 kDa MBP (Figure 2h, 2i), indicate a similar localziation of these two proteins in rat brain tissue. The signals of MBPs are enhanced in midbrain and white matter regions of the brain. MBPs are abundant proteins in the myelin sheath essential for the formation of myelin and associated with protein transport and signaling.[35][36][37]
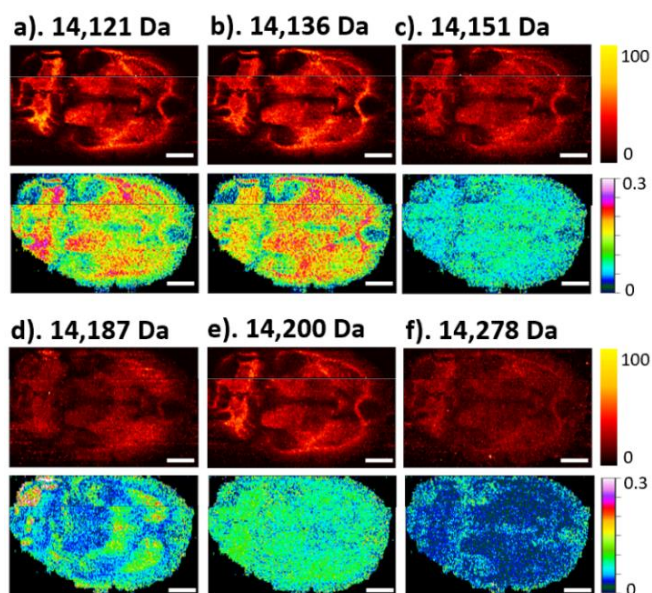

Figure 3. Ion images of the +19 charge state of the 14.1 kDa MBP proteoforms. Top panels show ion images normalized to the TIC. Bottom panels show images generated by plotting the ratio of the individual proteoform signal to the sum of signals of all the MBP proteoforms: (a) m/z 744.2335[19+],14,121Da, N-terminal acetylation, (b) m/z 745.0235[19+], 14,136Da, N-terminal acetylation, methionine sulfoxide, (c) m/z 745.8132[19+], 14,151 Da, phosphorylation, (d) m/z 747.7063[19+], 14,187 Da, N-terminal acetylation, phosphorylation, and -16 Da mass shift, (e) m/z 748.3901[19+], 14,200Da, N-terminal acetylation, S118 or S122 phosphorylation, (f) m/z 752.4923[19+], 14,278 Da, N-terminal acetylation, di-phosphorylation. Scale bar: 3mm.

MBP and hemoglobin are the two proteins with most PTMs detected in this study. A region of the mass spectrum containing the +19 charge state of 14.1 kDa MBP proteoforms and +25 charge state of 18.4 kDa MBP proteoforms is shown in Figure S6. Figure 3 shows ion images of six MBP proteoforms. Ion images of all the eleven MBP proteoforms are shown in Figure S7. Additional ion images obtained in replicate experiments are shown in Figure S8 and S9 to demonstrate reproducibility. All the proteoforms have similar spatial distributions, which is consistent with the expression of their coding genes in specific regions. However, the relative abundances of different proteoforms vary across the tissue. These differences are visualized using ratio images shown under each ion image in Figure 3. The ratio images were generated by plotting the ratio of the individual proteoform signal to the sum of signals of all the MBP proteoforms in the +19 charge state in each pixel of the image. The ratio images reveal the expression of each proteoform in different regions of the brain tissue.

Both the N-terminal acetylated MBP (Figure 3a) and its oxidized form (Figure 3b) are enhanced in the midbrain and white matter regions. However, the N-terminal acetylated MBP (Figure 3a) is more abundant in the white matter of the cerebellum region whereas the oxidized form is more abundant in the midbrain region (Figure 3b). Methionine oxidation is a spontaneous PTM attributed to the oxidative damage, which accumulates with aging.[38] Methionine sulfoxide reductase (MSR) repairs the oxidative damage by reducing methionine sulfoxide to methionine.[38] This observed difference in the relative abundance of the two proteoforms may be attributed to the higher activity of MSR in the cerebellum region in comparison with other regions of the brain.[39]

The ratio images of the phosphorylated MBP (Figure 3c) and its N-terminal acetylated form (Figure 3e) display a uniform distribution across the entire tissue section. However, the ratio image of the N-terminal acetylated, di-phosphorylated MBP (Figure 3f) shows the enhanced abundance of this proteoform in the hippocampal formation and isocortex regions. An interesting ratio image showing an enhanced abundance in the thalamus and striatum ventral region was observed for the N-terminal acetylated, phosphorylated and uncharacterized -16 Da mass shift proteoform (Figure 3d).

In summary, the combination of on-tissue top-down proteomics with nano-DESI MSI has been used to generate first proteoform-selective images of a rat brain tissue. In this proof-of-concept study, we obtained spatial maps of forty proteoforms of nineteen distinct proteins. Ratio images reveal the differential expression of the individual proteoforms in different parts of the tissue. The results obtained in this study show that direct imaging of proteoforms, with their assignment and identification, is feasible. With proteoform-selective imaging, the mapping of differentially modified proteoforms to functional units within tissues is possible, which will help in the assignment of their function. As the technology is pushed toward the single cell level, the exploration of proteoforms in specific cell types in human tissues will be possible, thus advancing our understanding of proteoform function and their exploitation as disease biomarkers.

[1] R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. Maccoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young, B. Zhang, *Nat. Chem. Biol.* **2018**, *14*, 206–214.

[2] L. M. Smith, N. L. Kelleher, *Nat. Methods* **2013**, *10*, 186–187.

[3] R. B. Parekh, C. Rohlff, *Curr. Opin. Biotechnol.* **1997**, *8*, 718–723.

[4] N. Siuti, N. L. Kelleher, *Nat. Methods* **2007**, *4*, 817–821.

[5] E. N. Mccool, R. A. Lubeckyj, X. Shen, D. Chen, Q. Kou, X. Liu, L. Sun, *Anal. Chem.* **2018**, *90*, 5529–5533.

[6] X. Shen, Z. Yang, E. N. McCool, R. A. Lubeckyj, D. Chen, L. Sun, *TrAC - Trends Anal. Chem.* **2019**, *120*, 115644.

[7] J. C. Tran, L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. Li, C. Wu, S. M. M. Sweet, B. P. Early, N. Siuti, R. D. Leduc, P. D. Compton, P. M. Thomas, N. L. Kelleher, *Nature* **2011**, *480*, 254–258.

[8] L. C. Anderson, C. J. Dehart, N. K. Kaiser, R. T. Fellers, D. F. Smith, J. B. Greer, R. D. Leduc, G. T. Blakney, P. M. Thomas, N. L. Kelleher, C. L. Hendrickson, *J. Proteome Res.* **2016**, *16*, 1087–1096.

[9] Y. Dai, K. E. Buxton, L. V Schaffer, R. M. Miller, R. J. Millikin, M. Scalf, B. L. Frey, M. R. Shortreed, L. M. Smith, *J. Proteome Res.* **2019**, *18*, 3671–3680.

[10] J. Park, P. D. Piehowski, C. Wilkins, M. Zhou, J. Mendoza, G. M. Fujimoto, B. C. Gibbons, J. B. Shaw, Y. Shen, A. K. Shukla, R. J. Moore, T. Liu, V. A. Petyuk, N. Tolić, L. Paša-Tolić, R. D. Smith, S. H. Payne, S. Kim, *Nat. Methods* **2017**, *14*, 909–914.

[11] L. Smith, J. Agar, J. Chamot-Rooke, P. Danis, Y. Ge, J. Loo, L. Pasa-Tolic, Y. Tsybin, N. Kelleher, **2020**, DOI 10.20944/PREPRINTS202010.0368.V1.

[12] I. Ntai, L. Fornelli, C. J. Dehart, J. E. Hutton, P. F. Doubleday, R. D. Leduc, A. J. Van Nispen, R. T. Fellers, G. Whiteley, E. S. Boja, H. Rodriguez, N. L. Kelleher, *PNAS* **2018**, *115*, 4140–4145.

[13] J. Zhang, M. J. Guy, H. S. Norman, Y.-C. Chen, Q. Xu, X. Dong, H. Guner, S. Wang, T. Kohmoto, K. H. Young, R. L. Moss, Y. Ge, H. Proteomics Program, *J. Proteome Res* **2011**, *10*, 4054–4065.

[14] B. M. Prentice, D. J. Ryan, R. Van De Plas, R. M. Caprioli, J. M. Spraggins, *Anal. Chem.* **2018**, *90*, 5090–5099.

[15] R. M. Caprioli, *J. Biomol. Tech.* **2019**, *30*, 7–11.

[16] A. Zavalin, J. Yang, K. Hayden, M. Vestal, R. M. Caprioli, *Anal Bioanal Chem* **2015**, *407*, 2337–2342.

[17] D. J. Ryan, N. H. Patterson, N. E. Putnam, A. D. Wilde, A. Weiss, W. J. Perry, J. E. Cassat, E. P. Skaar, R. M. Caprioli, J. M. Spraggins, *Anal. Chem.* **2019**, *91*, 7578–7585.

[18] B. Cillero-Pastor, R. M. A. Heeren, *J. Proteome Res.* **2013**, *13*, 325–335.

[19] Y. Zhu, P. D. Piehowski, R. Zhao, J. Chen, Y. Shen, R. J. Moore, A. K. Shukla, V. A. Petyuk, M. Campbell-Thompson, C. E. Mathews, R. D. Smith, W.-J. Qian, R. T. Kelly, *Nat. Commun.* **2018**, *9*, 882.

[20] P. D. Piehowski, Y. Zhu, L. M. Bramer, K. G. Stratton, R. Zhao, D. J. Orton, R. J. Moore, J. Yuan, H. D. Mitchell, Y. Gao, B. J. M. Webb-Robertson, S. K. Dey, R. T. Kelly, K. E. Burnum-Johnson, *Nat. Commun.* **2020**, *11*, 1–12.

[21] R. L. Griffiths, K. I. Kocurek, H. J. Cooper, *Curr. Opin. Chem. Biol.* **2018**, *42*, 67–75.

[22] M. W. Towers, T. Karancsi, E. A. Jones, S. D. Pringle, E. Claude, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2456–2466.

[23] K. Y. Garza, C. L. Feider, D. R. Klein, J. A. Rosenberg, J. S. Brodbelt, L. S. Eberlin, *Anal. Chem.* **2018**, *90*, 7785–7789.

[24] O. J. Hale, E. K. Sisley, R. L. Griffiths, I. B. Styles, H. J. Cooper, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 873–879.

[25] O. J. Hale, H. J. Cooper, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2531–2537.

[26] R. Yin, K. E. Burnum-Johnson, X. Sun, S. K. Dey, J. Laskin, *Nat. Protoc.* **2019**, *14*, 3445–3470.

[27] P. J. Roach, J. Laskin, A. Laskin, *Analyst* **2010**, *135*, 2161–2452.

[28] C. C. Hsu, P. T. Chou, R. N. Zare, *Anal. Chem.* **2015**, *87*, 11171–11175.

[29] C.-L. Chen, T.-H. Kuo, H.-H. Chung, P. Huang, L.-E. Lin, C.-C. Hsu, *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 653–660.

[30] O. J. Hale, H. J. Cooper, *Anal. Chem.* **2021**, *93*, 4619–4627.

[31] L. M. Smith, P. M. Thomas, M. R. Shortreed, L. V Schaffer, R. T. Fellers, R. D. LeDuc, T. Tucholski, Y. Ge, J. N. Agar, L. C. Anderson, J. Chamot-Rooke, J. Gault, J. A. Loo, L. Paša-Tolić, C. V Robinson, H. Schlüter, Y. O. Tsybin, M. Vilaseca, J. Antonio Vizcaíno, P. O. Danis, N. L. Kelleher, *Nat. Methods* **2019**, *16*, 939–940.

[32] R. E. Gossett, A. A. Frolov, J. B. Roths, W. David Behnke, A. B. Kier, F. Schroeder B', *Lipids* **1996**, *31*, 895–918.

[33] P. Carpintero, R. Anado´n, A. Anado´n, S. Di´az, D. Di´az-Regueira, J. Go´mez, G. Go´mez-Ma´rquez, M. Ma´rquez, *Neuroscience* **1999**, *90*, 1433–1444.

[34] G. H. Zhang, K. D. Murthy, R. Binti Pare, Y. H. Qian, *Eur. J. Inflamm.* **2020**, *18*, 1–11.

[35] D. A. Plymire, C. E. Wing, D. E. Robinson, S. M. Patrie, *Anal. Chem.* **2017**, *89*, 12030–12038.

[36] *Cell. Mol. Life Sci.* **2006**, *63*, 1945–1961.

[37] J. M. Boggs, *Cell. Mol. Life Sci.* **2006**, *63*, 1945–1961.

[38] P. A. C. Cloos, S. Christgau, *Biogerontology* **2004**, *5*, 139–158.

[39] E. R. Stadtman, H. Van Remmen, A. Richardson, N. B. Wehr, R. L. Levine, *Biochim. Biophys. Acta - Proteins Proteomics* **2005**, *1703*,

135–140.