Neutral loss mass spectral data enhances molecular similarity analysis in METLIN

Aisporna, A.; Benton, H.P.; Chen, A.; Derks, R.J.E.; Galano, J.M.; Giera, M.; Siuzdak, G.

Research Article

Neutral Loss Mass Spectral Data Enhances Molecular Similarity Analysis in METLIN

Aries Aisporna, H. Paul Benton, Andy Chen, Rico J. E. Derks, Jean Marie Galano, Martin Giera, and Gary Siuzdak*

Cite This: *J. Am. Soc. Mass Spectrom.* 2022, 33, 530−534

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Neutral loss (NL) spectral data presents a mirror of MS[2] data and is a valuable yet largely untapped resource for molecular discovery and similarity analysis. Tandem mass spectrometry (MS[2]) data is effective for the identification of known molecules and the putative identification of novel, previously uncharacterized molecules (unknowns). Yet, MS[2] data alone is limited in characterizing structurally related molecules. To facilitate unknown identification and complement the METLIN-MS[2] fragment ion database for characterizing structurally related molecules, we have created a MS[2] to NL converter as a part of the METLIN platform. The converter has been used to transform METLIN's MS[2] data into a neutral loss database (METLIN-NL) on over 860 000 individual molecular standards. The platform includes both the MS[2] to NL converter and a graphical user interface enabling comparative analyses between MS[2] and NL data. Examples of NL spectral data are shown with oxylipin analogues and two structurally related statin molecules to demonstrate NL spectra and their ability to help characterize structural similarity. Mirroring MS[2] data to generate NL spectral data offers a unique dimension for chemical and metabolite structure characterization.
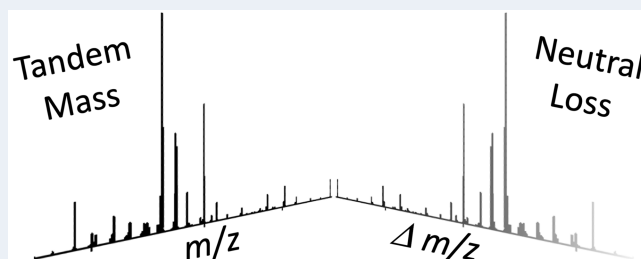
Similarity analysis[1−4] and molecular networking[5,6] using tandem mass spectrometry (MS[2]) data have become valuable approaches for identifying previously uncharacterized molecules (unknowns).[1] Yet key structural information can be lost when relying solely on fragment ion data, for example, the loss of a sulfate ion from two similar molecules of different masses will not result in fragment ion overlap.[7] This is of significant practical relevance. A user who would try to identify an unknown based on a MS[2] database similarity search would not succeed in obtaining structurally relevant matches. However, retrieving this structurally useful information is possible by analyzing the differences between the molecular ion and the fragment ion, or better known as the neutral loss (NL) and symbolized by $\Delta m/z$. NLs[1,2] constitute a rich resource and have already been widely used in proteomics, pharmacology, and metabolomics for over three decades[1,2,8−12] as represented by over a thousand papers on the topic. Yet, even though mass spectrometry-based NL analysis has been extensively applied, no small molecule MS[2] to NL conversion programs exist, nor any comprehensive library of NL spectra.

Unlike MS[2] data, which projects the m/z values and intensity of the precursor and each fragment, for NL data $\Delta m/z$, is projected as the difference between precursor ions and its respective fragment ions. It also can be generated as a difference between fragment ions.[10]

The new METLIN NL converter has been created as a general resource and to convert METLIN's 860 000 MS[2] small

molecule molecular standards database into a mirrored NL database (METLIN-NL) to facilitate neutral loss searching. The NL data was derived across a broad range of standards representing hundreds of different chemical classes.[3,13] The converter was designed to input METLIN's MS[2] data and convert it to METLIN-NL spectra (e.g., Figure 1 asymmetric dimethylarginine (ADMA)) by using the converter to calculate the differences between the precursor molecular ion and the fragment ions in the experimental MS[2] mass spectra (Figure 1A). The NL spectra ($NL_{intensity}$ versus $\Delta m/z$) were created (e.g., ADMA Figure 1B) with the NL intensity ($NL_{intensity}$) using the fragment ion intensities from each precursor/fragment generated NL ($\Delta m/z$). It should be noted that not all precursor to fragment peaks represent a true NL between the precursor and fragment ions, and therefore some of the peaks in the NL spectra can also be considered (as recently described[10]) hypothetical neutral losses.
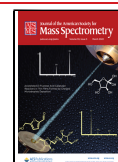
The MS[2] to NL converter (https://metlin.scripps.edu and https://github.com/masspec/MS2ToNLConverter) allows
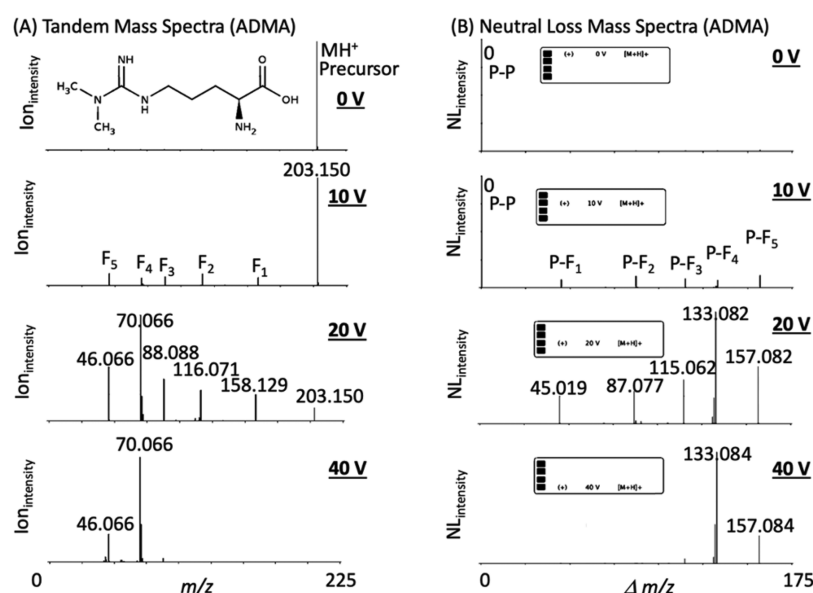
530

Figure 1. METLIN-NL mass spectral database was derived from the METLIN-MS^2 data on over 860 000 molecular standards. (A) Asymmetric dimethylarginine (ADMA) and its representative METLIN-MS^2 spectra at four different collision energies. (B) METLIN-NL spectra (NL$_{intensity}$ versus $\Delta m/z$) of ADMA was generated by calculating the difference between the precursor and fragment ions with NL$_{intensity}$ based on the original fragment ion intensities. "P" refers to precursor ion and "F" refers to fragment ion.

users to view a single MS^2 or NL spectra or do a comparison between two MS^2 or NL data.

When using METLIN IDs, the MS^2 and NL data are already calculated, but when using CSV files the MS^2 data is automatically converted to NL. To facilitate these analyses, METLIN-NL is built on a Linux platform with the initial version of the graphical user interface (GUI) created using Highcharts, HTML, JQuery, MySQL, and PHP. The GUI allows for comparative analyses between different compounds including neutral loss data (NL$_{int}$ versus $\Delta m/z$) as well as MS/MS data (Frag$_{int}$ versus m/z) in both positive and negative ionization modes. The GU also offers visualization either at each individual collision energy, or a "composite spectra", that constitutes all spectra across the multiple collision energies. Once a spectrum (or spectra) has been generated, users will be able to hover the cursor on each peak to obtain detailed information about m/z, intensity, ionization mode, the compound's name, and collisional energy values. The user input, for example, CSV file, for the Web site https://metlin-nl. scripps.edu/ requires a compound name, masses with intensities, collision energy, positive/negative mode, and precursor value. Users have access to two downloadable CSV file to demonstrate formatting.

The converter operates in the following modes and allows users to create/compare the following data types (Table 1).

Table 1. MS^2 to NL Spectra Converter and Its Modes of Operation

input #1	input #2	graph type
METLIN ID		shows MS^2 and NL spectra
METLIN ID	METLIN ID	compares MS^2 and NL spectra
CSV		shows MS^2 and NL spectra
CSV	CSV	compares MS^2 and NL spectra
METLIN ID	CSV	compares MS^2 and NL spectra
CSV	METLIN ID	compares MS^2 and NL spectra

METLIN-NL is a compilation of NL$_{intensity}$ versus $\Delta m/z$ spectra generated from METLIN's eight distinct MS^2 data sets created from 860 000 standards.[3] This compilation is represented within METLIN-NL at four different collision energies and in both positive and negative ionization modes. The rationale behind providing multiple conditions is that MS^2 collision energies have not been standardized and such broad acquisition parameters are required to represent the output across different instrument types. An additional rationale for the array of conditions is that different molecules can fragment differently depending on the collision energies thus METLIN provides a broad range of empirical data across its 860 000 standards. It is worth noting that all of METLIN's MS^2 data is empirical data and has not been generated from predictive *in silico*-based approaches.

A secondary set of METLIN-NL data has also been accumulated based on precursor minus fragment ion transitions as well as all possible fragment-to-fragment ion transitions to provide a more comprehensive set of experimentally derived structural data. Unlike the original METLIN-MS^2 database, METLIN-NL represents a translation that more effectively enables the molecular annotation of unknown molecular entities since NL data inherently corrects for molecular weight differences.

To test the utility of METLIN-NL we examined two different types of molecular structures, oxylipins and a pharmaceutical (statin) drug and its demethylated metabolite. Oxylipins[14] represent a class of highly active lipid metabolites ubiquitous in humans and plants, and specifically the phytoprostanes (PhytoPs) class of oxylipins resemble prostaglandin-like compounds that are found in seeds and vegetable oils derived from oxidative cyclization of α-linolenic acid. Since PhytoPs are a class of highly structurally related oxylipins and are suspected to have additional unidentified analogs,[14−16] we chose them to demonstrate the utility of METLIN-NL. Tandem MS and NL data were recently generated on a set of PhytoPs, including the structural analogs
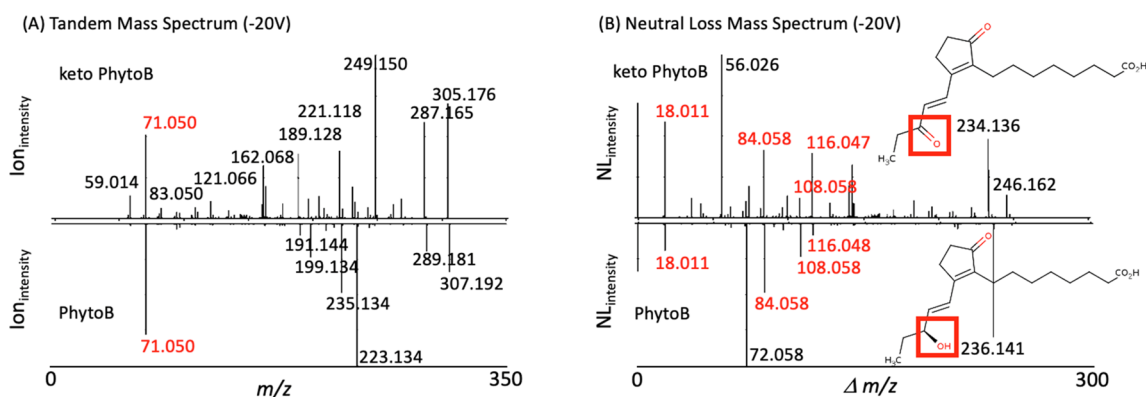
Figure 2. MS^2 and NL data on two related oxylipins (16 keto 16-B$_1$-PhytoP and 16-B$_1$-PhytoP). (A) Oxylipin MS^2 data show little overlap (in red) in contrast to the (B) NL spectra with the high-resolution neutral loss data facilitating similarity analysis with both providing complementary structural information. The red box denotes the only (minor) structural differences between the two molecules.
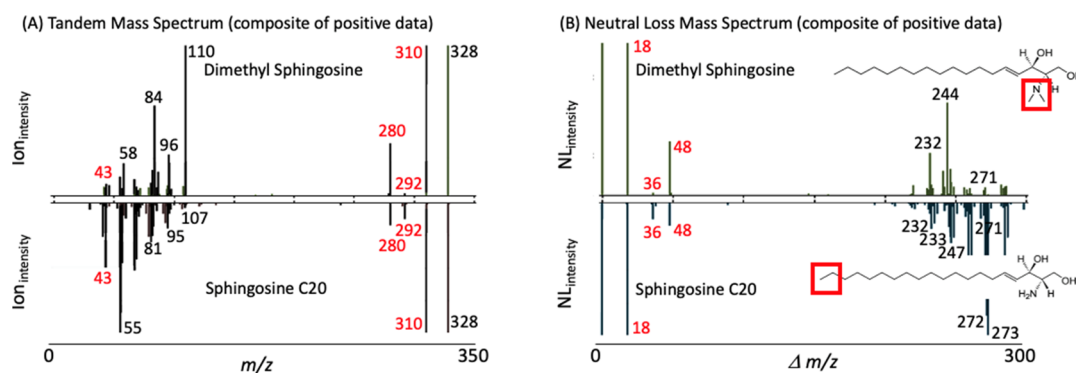

Figure 3. Synergy between MS^2 and neutral loss data on two related sphingosine molecules. Dimethyl sphingosine and sphingosine C20 (A) MS^2 and (B) neutral loss data; both the MS^2 data and the neutral loss data show both overlapping and distinctly different peaks. The red box denotes the structural differences between the two molecules.
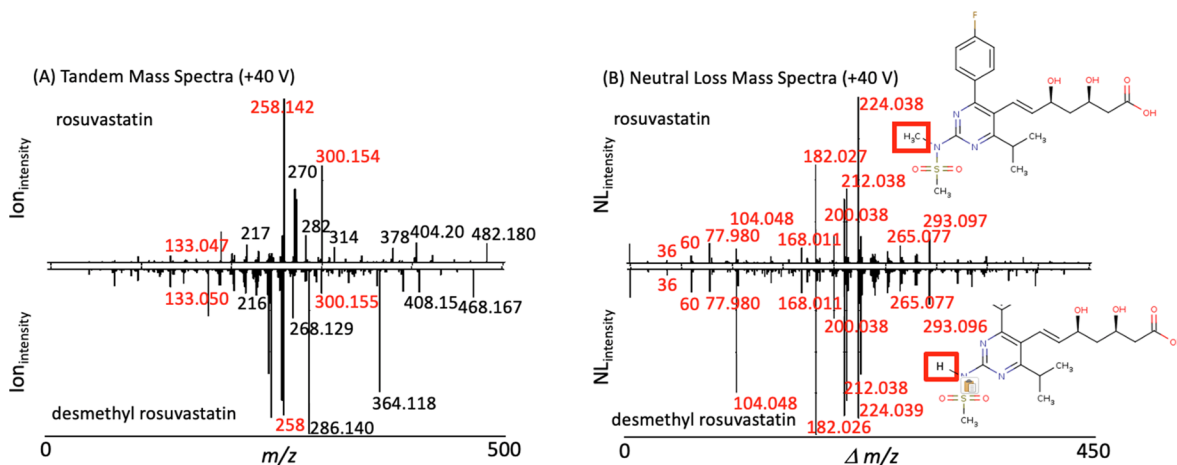

Figure 4. MS^2 and NL data on two related statins. Rosuvastatin and desmethyl rosuvastin (A) MS^2 and (B) NL data; the MS^2 data show few overlapping peaks (in red) while the NL spectra provide near complete overlap. Interestingly, while the NLs help facilitate similarity, the MS^2 data provides more structurally distinguishing features. The red box denotes a minor structural difference between the two molecules.

16-B$_1$-PhytoP and 16-keto 16-B$_1$-PhytoP (Figure 2). When trying to extrapolate/correlate the observed tandem MS spectra of the two PhytoPs, classic similarity searching was of very limited value providing only one overlapping ion, even though some fragments presented an expected two Dalton difference (Figure 2A). This exemplifies that two structurally very similar molecules can yield highly different MS^2 spectra limiting similarity searching possibilities and thereby severely

impacting the usefulness of this approach for the identification of chemically closely related substances. However, NL similarity analysis yielded multiple overlapping NLs (Figure 2B). Further analysis of the tandem MS data as well as the molecular weight difference between the two molecules being 2 Da were consistent with 16-keto 16-B1-PhytoP. This NL data (unlike the MS^2 data) helped to easily correlate the two molecules, and the distinguishing NL and fragment ions

exclusive to 16-keto 16-B$_1$-PhytoP and 16-B$_1$-Phyto provided significant structural information.

Another example with dimethyl sphingosine and sphingosine C20 (Figure 3A,B) further shows the synergy that MS^2 and NL data can have between structural analogues. Dimethyl sphingosine and sphingosine C20 (A) MS^2 and (B) NL data each have the same elemental composition yet distinct structures. In this case, both the MS^2 data and the NL data show multiple overlapping peaks thus representing an example where both types of complementary data provide confirmation of the structural similarity yet each has unique distinguishing information.

The purpose of having a large database is to help reduce the need for speculation and allow for the rapid identification of molecules. However, since many molecular structures are not represented in any database, similarity analyses offer an alternative in the preliminary characterization process. This process extends beyond naturally occurring molecules and can be applied just as readily to xenobiotics and other chemical entities. The third example in applying METLIN-NL is shown here for a nonendogenous drug molecule and its metabolite.

The well-known cholesterol-lowering statin drug rosuvastatin[17] (trade name Crestor) and its active metabolite desmethyl rosuvastatin[18] differ in mass by 14 Da (demethylation reaction) and the MS^2 and NL data (Figure 4A,B) of these two molecules have recently been acquired and populated within METLIN and METLIN-NL. As was observed with the oxylipins, tandem MS data was of limited utility when searching METLIN (Figure 4A), where three fragment ions were overlapping between the two molecules. However, NL matching/detection showed near complete overlap (Figure 4B). Further analysis of the tandem MS data as well as the molecular weight difference between the two molecules being 14 Da were consistent with loss of a methyl group. For the rosuvastatin NL data, the overlap in the NL data clearly dominated the comparative analyses, making similarity searching much more effective using NL while the MS^2 data provided complementary information that was informative for structural determination. Overall, the NL data, which were completely derived from the MS^2 data, are more effective (than MS^2) at showing similarity.

METLIN's molecular standards with systematically acquired experimental MS^2 data across multiple collision energies, allows for the comprehensive generation and graphical user interface (beta) visualization (Figure 4) of NL data. Fragment ion and NL similarity analysis[1] was originally developed to aid in the identification of novel molecules (unknowns)[1] by using fragment ion and NL data to help align an unknown molecule to compounds with similar fragmentation data within a database. However, now with a NL database of small molecules available via METLIN-NL NL similarity analysis can be more readily applied to a host of biological and chemical challenges.

Overall, The METLIN MS^2 to NL converter and METLIN-NL empirically derived data will enable new types of analyses facilitating more rapid identification of unknown compounds via both fragment ion and NL similarity searching.[2] Both biologists and chemists will be able to apply METLIN-NL to the structure elucidation of unknowns derived from animals,[19] plants,[14,20] or microbiota;[21] and METLIN-NL can also be used as a resource for informatics[22] as well as identifying unexpected synthetic chemical or enzymatically modified drug products (e.g., pharmaceuticals[23]) as it is populated with both biological and chemical entities. Given METLIN's extensive userbase[3]

and the ubiquitous application of mass spectrometry-based NL analysis (dating back three decades), METLIN-NL and its MS^2 to NL converter promises to have wide-ranging utility.

■ AUTHOR INFORMATION

Corresponding Author

Gary Siuzdak − *Scripps Center for Metabolomics and Mass Spectrometry and Department of Chemistry, Molecular and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, United States;* ⊙ orcid.org/0000-0002-4749-0014; Phone: 858 784 9425; Email: siuzdak@scripps.edu

Authors

Aries Aisporna − *Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, California 92037, United States*

H. Paul Benton − *Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, California 92037, United States*

Andy Chen − *Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, California 92037, United States*

Rico J. E. Derks − *Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333ZA Leiden, Netherlands;* ⊙ orcid.org/0000-0002-8920-7133

Jean Marie Galano − *Institut des Biomolécules Max Mousseron, UMR 5247 CNRS, ENSCM, Université de Montpellier, 34000 Montpellier, France;* ⊙ orcid.org/0000-0001-6412-4967

Martin Giera − *Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333ZA Leiden, Netherlands;* ⊙ orcid.org/0000-0003-1684-1894

Complete contact information is available at:
https://pubs.acs.org/10.1021/jasms.1c00343

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. XCMS^2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. *Anal. Chem.* **2008**, *80*, 6382−6389.

(2) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. METLIN: A technology platform for identifying knowns and unknowns. *Anal. Chem.* **2018**, *90* (5), 3156−64.

(3) Xue, J.; Guijas, C.; Benton, H. P.; Warth, B.; Siuzdak, G. METLIN MS^2 molecular standards database: a broad chemical and biological resource. *Nat. Methods* **2020**, *17*, 953−954.

(4) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat. Biotechnol.* **2012**, *30*, 826−828.

(5) Li, D.; Baldwin, I. T.; Gaquerel, E. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (30), E4147−4155.

(6) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (26), E1743−E1752.

(7) Flasch, M.; Bueschl, C.; Woelflingseder, L.; Schwartz-Zimmermann, H. E.; Adam, G.; Schuhmacher, R.; Marko, D.; Warth, B. Stable Isotope-Assisted Metabolomics for Deciphering Xenobiotic Metabolism in Mammalian Cell Culture. *ACS Chem. Biol.* **2020**, *15*, 970−981.

(8) Martin, D. B.; Eng, J. K.; Nesvizhskii, A. I.; Gemmill, A.; Aebersold, R. Investigation of neutral loss during collision-induced dissociation of peptide ions. *Anal. Chem.* **2005**, *77*, 4870−4882.

(9) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703−714.

(10) Xing, S.; Hu, Y.; Yin, Z.; Liu, M.; Tang, X.; Fang, M.; Huan, T. Retrieving and Utilizing Hypothetical Neutral Losses from Tandem Mass Spectra for Spectral Similarity Analysis and Unknown Metabolite Annotation. *Anal. Chem.* **2020**, *92*, 14476−14483.

(11) Heller, D. N.; Murphy, C. M.; Cotter, R. J.; Fenselau, C.; Uy, O. M. Constant neutral loss scanning for the characterization of bacterial phospholipids desorbed by fast atom bombardment. *Anal. Chem.* **1988**, *60*, 2787−2791.

(12) Schwudke, D.; Oegema, J.; Burton, L.; Entchev, E.; Hannich, J. T.; Ejsing, C. S.; Kurzchalia, T.; Shevchenko, A. Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. *Anal. Chem.* **2006**, *78*, 585−595.

(13) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **2016**, *8*, 61.

(14) Galano, J. M.; Lee, Y. Y.; Oger, C.; Vigor, C.; Vercauteren, J.; Durand, T.; Giera, M.; Lee, J. C. Isoprostanes, neuroprostanes and phytoprostanes: An overview of 25 years of research in chemistry and biology. *Prog. Lipid Res.* **2017**, *68*, 83−108.

(15) Watrous, J. D.; Niiranen, T. J.; Lagerborg, K. A.; Henglin, M.; Xu, Y. J.; Rong, J.; Sharma, S.; Vasan, R. S.; Larson, M. G.; Armando, A.; Mora, S.; Quehenberger, O.; Dennis, E. A.; Cheng, S.; Jain, M. Directed Non-targeted Mass Spectrometry and Chemical Networking for Discovery of Eicosanoids and Related Oxylipins. *Cell Chem. Biol.* **2019**, *26* (3), 433−442.

(16) Young, R. S. E.; Bowman, A. P.; Williams, E. D.; Tousignant, K. D.; Bidgood, C. L.; Narreddula, V. R.; Gupta, R.; Marshall, D. L.; Poad, B. L. J.; Nelson, C. C.; Ellis, S. R.; Heeren, R. M. A.; Sadowski, M. C.; Blanksby, S. J. Apocryphal FADS2 activity promotes fatty acid diversification in cancer. *Cell Rep.* **2021**, *34*, 108738.

(17) Fellström, B. C.; Jardine, A. G.; Schmieder, R. E.; Holdaas, H.; Bannister, K.; Beutler, J.; Chae, D. W.; Chevaile, A.; Cobbe, S. M.; Grönhagen-Riska, C.; De Lima, J. J.; Lins, R.; Mayer, G.; McMahon, A. W.; Parving, H. H.; Remuzzi, G.; Samuelsson, O.; Sonkodi, S.; Sci, D.; Süleymanlar, G.; Tsakiris, D.; Tesar, V.; Todorov, V.; Wiecek, A.; Wüthrich, R. P.; Gottlow, M.; Johnsson, E.; Zannad, F. AURORA Study Group. Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. *N Engl J. Med.* **2009**, *360*, 1395−1407.

(18) Martin, P. D.; Warwick, M. J.; Dane, A. L.; Hill, S. J.; Giles, P. B.; Phillips, P. J.; Lenz, E. Metabolism excretion, and pharmacokinetics of rosuvastatin in healthy adult male volunteers. *Clin Ther.* **2003**, *25* (11), 2822−2835.

(19) Rosenberg, G.; Yehezkel, D.; Hoffman, D.; Mattioli, C. C.; Fremder, M.; Ben-Arosh, H.; Vainman, L.; Nissani, N.; Hen-Avivi, S.; Brenner, S.; Itkin, M.; Malitsky, S.; Ohana, E.; Ben-Moshe, N. B.; Avraham, R. Host succinate is an activation signal for *Salmonella* virulence during intracellular infection. *Science* **2021**, *371* (6527), 400−405.

(20) Lipan, L.; Collado-González, J.; Domínguez-Perles, R.; Corell, M.; Bultel-Poncé, V.; Galano, J. M.; Durand, T.; Medina, S.; Gil-Izquierdo, Á.; Carbonell-Barrachina, Á. Phytoprostanes and Phytofurans-Oxidative Stress and Bioactive Compounds-in Almonds are Affected by Deficit Irrigation in Almond Trees. *J. Agric. Food Chem.* **2020**, *68* (27), 7214−7225.

(21) Guo, H.; Chou, W. C.; Lai, Y.; Liang, K.; Tam, J. W.; Brickey, W. J.; Chen, L.; Montgomery, N. D.; Li, X.; Bohannon, L. M.; Sung, A. D.; Chao, N. J.; Peled, J. U.; Gomes, A. L. C.; van den Brink, M. R. M.; French, M. J.; Macintyre, A. N.; Sempowski, G. D.; Tan, X.; Sartor, R. B.; Lu, K.; Ting, J. P. Y. Multi-omics analyses of radiation survivors identify radioprotective microbes and metabolites. *Science* **2020**, *370*, 6516.

(22) Wandy, J.; Zhu, Y.; van der Hooft, J. J.; Daly, R.; Barrett, M. P.; Rogers, S. Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics* **2018**, *34* (2), 317−318.

(23) Giera, M.; de Vlieger, J. S.; Lingeman, H.; Irth, H.; Niessen, W. M. Structural elucidation of biologically active neomycin N-octyl derivatives in a regioisomeric mixture by means of liquid chromatography/ion trap time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24* (10), 1439−1446.