PERSPECTIVE OPEN ACCESS

The Dark Metabolome/Lipidome and In-Source Fragmentation

Winnie Uritboonthai[1] | Linh Hoang[1] | Aries Aisporna[1] | Martin Giera[2,3] | Gary Siuzdak[1,4]

[1]Scripps Center of Metabolomics and Mass Spectrometry, La Jolla, California, USA | [2]Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, The Netherlands | [3]The Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW), Leiden University Medical Center, Leiden, The Netherlands | [4]Department of Chemistry, Molecular and Computational Biology Scripps Research, La Jolla, California, USA

Correspondence: Gary Siuzdak (siuzdak@scripps.edu)

To the editor,

Tandem mass spectrometry (MS/MS) is valued for its ability to facilitate molecular identification and deliver highly consistent data across a wide range of mass spectrometry platforms. Distinct from MS/MS is the fragmentation that occurs during electrospray ionization (ESI), commonly referred to as in-source fragmentation (ISF) (Figure 1). ISF was first observed in the 1950s with electron ionization and has been recognized as an inherent yet often overlooked feature of the ESI process, albeit less prevalent than with electron ionization. Recently, ISF has been associated with the overrepresentation of peaks in liquid chromatography mass spectrometry (LC/MS) data, where it accounts for the majority of observed unfiltered peaks [1]. Due to its overrepresentation in LC/MS data, and the subsequent inability to identify the molecules associated with these peaks using MS/MS data, ISF has been linked to the so-called "dark metabolome" [2, 3] (also encompassing the lipidome), a term used to describe uncharacterized molecular species in metabolomics and lipidomics. This association [1] was determined by an examination of MS/MS data acquired at 0 eV collision energy from METLIN's extensive library of over 931,000 molecular standards. However, while the similarity of ISF and MS/MS at 0 eV data has been described in previous studies [1, 4–6], it has yet to be directly established that they correlate with each other. We explored the consistency between MS/MS (0 eV) data and ISF across various molecular species to assess whether mining METLIN's MS/MS (0 eV) data—comprising over 931,000 molecular standards—can effectively link ISF to the dark metabolome and lipidome.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) with ESI has become a cornerstone in metabolomics, lipidomics, and clinical analysis due to its accuracy in identifying small molecules within complex biological matrices. With LC-MS/MS, after ionization occurs in the ESI source, charged molecules are directed into a collision cell where they undergo fragmentation for structural analysis. This procedure is typically repeated for all charged analytes present in a sample. However, despite its utility, this method has revealed an unexpectedly vast array of spectral features associated with the "dark metabolome." However, given the limited number of protein-coding genes [7, 8] with only a fraction producing enzymes, the chemical diversity [3, 9, 10] detected through LC-MS/MS—potentially hundreds of thousands or even millions of metabolites—far exceeds biological expectations. Current estimates suggest that less than 2% of observed LC-MS/MS spectra can be annotated, a potentially broad spectrum of unknown compounds [3]. Recent research [1] using the METLIN database and its data at 0 eV has shed light on this discrepancy, and much of the perceived complexity may stem from technological factors, particularly ISF, rather than from biological diversity itself.

Our laboratory, along with several others [11], has observed the widespread occurrence of ISF [12, 13]. This process involves the fragmentation of analytes during the initial ionization stage within the ESI source, occurring before they reach the collision
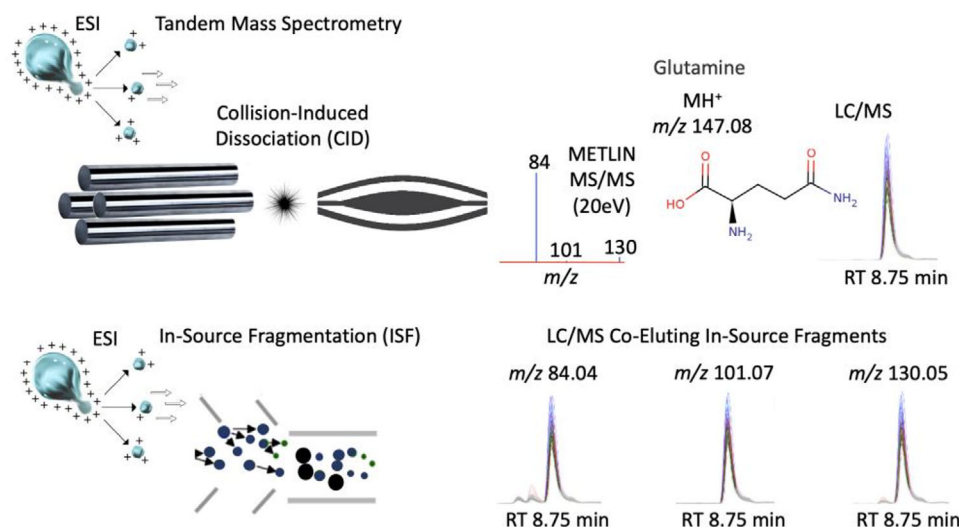
FIGURE 1 | Comparison of electrospray ionization (ESI) tandem mass spectrometry and in-source fragmentation (ISF) as separate events within a mass spectrometer. MS/MS results from collision-induced dissociation (CID) events between two mass analyzers, while ISF occurs in the ESI source. Example extracted ion chromatogram (EIC) data obtained from glutamine and the corresponding EIC from six distinct gnotobiotic sample groups. The consistency between the molecular ion and the ISF with respect to relative ion intensity and retention time provides a distinguishing feature of ISF.
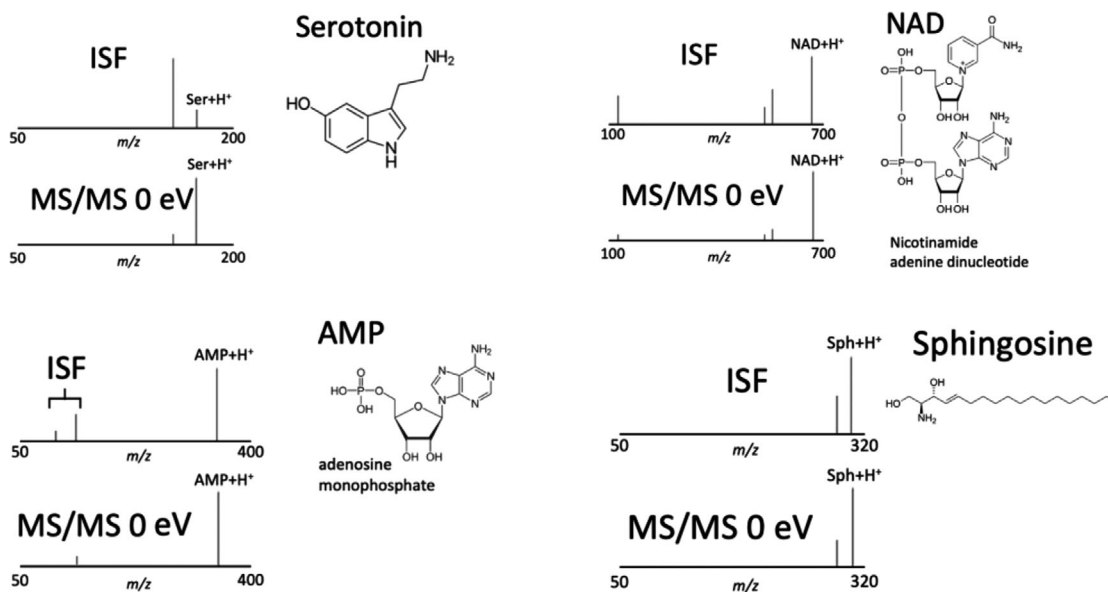

FIGURE 2 | Comparison of electrospray ionization (ESI) in-source fragmentation (ISF) and tandem mass spectrometry at 0 eV collision energy of endogenous metabolites/lipids. ISF is a common phenomenon that occurs in the ESI source and has been investigated here across common metabolite and lipid molecular standards, including serotonin, NAD, AMP, and sphingosine. The analysis reveals high consistency between ISF and tandem mass spectrometry (0 eV collisional energy).

cell. Essentially, ISF can transform a single analyte into multiple molecular ions and fragments, creating a complex array of ions from what was initially a single entity. Consequently, the mass analyzer indiscriminately isolates and further fragments whatever enters the collision cell. Given this understanding, we suspect that ISF may play a significant role in contributing to the so-called dark metabolome.

In order to correlate the observation of peaks and ISF, we examined the METLIN MS/MS database [14], which consists of over 931,000 molecular standards representing over 350 chemical classes in which we mined METLIN's MS/MS data at 0 eV, an energy designed to simulate the absence of CID. This analysis was performed to assess whether MS/MS spectra acquired at 0 eV collision energy in METLIN could reflect ISF-related fragments. The analysis revealed that ISF could account for over 70% of the peaks observed in typical LC-MS/MS metabolomic datasets when using a 5% cutoff threshold. This number rises when the threshold is reduced to less than 3%. The 5% and 3% thresholds represent a conservative range of peak intensities across LC/MS experiments, where the typical intensity count numbers range from 10000 to millions, well over two orders of magnitude.
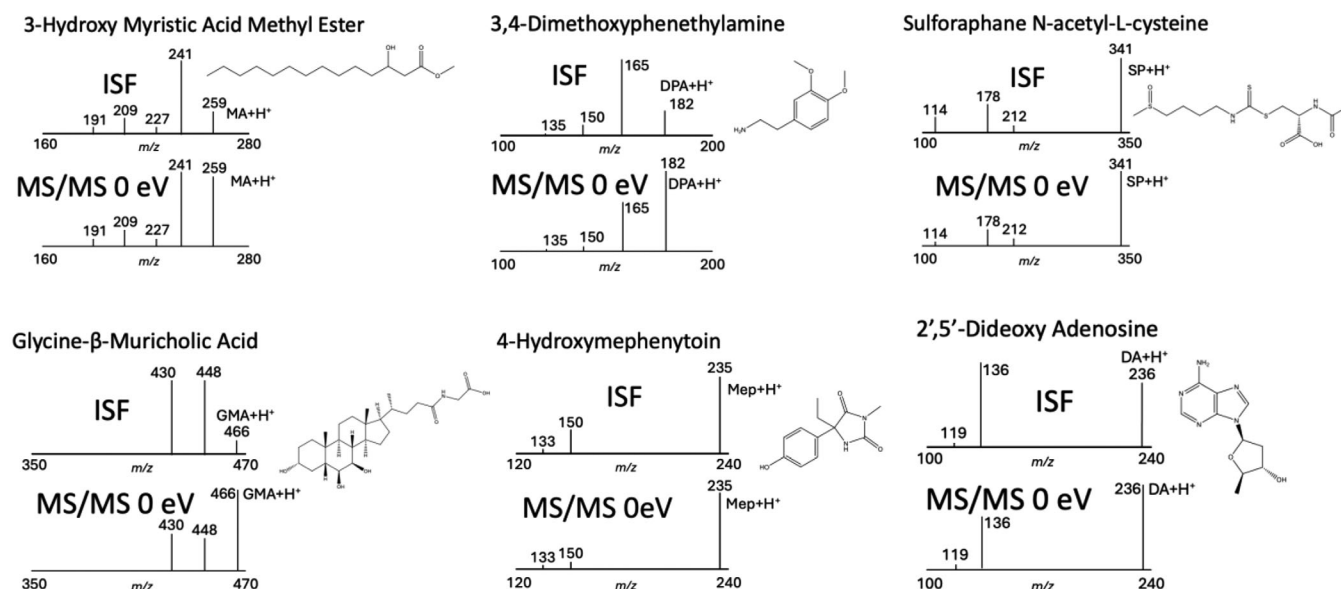
FIGURE 3 | Additional comparisons of ISF and MS/MS (0 eV) of nonbiological molecules, including common metabolite/lipid derivatives designed for therapeutic purposes. These include 3-hydroxy myristic acid methyl ester, 3,4-dimethoxyphenethylamine, sulforaphane N-acetyl-L-cysteine, glycine-β-muricholic acid, 4-hydroxymephenytoin, and 2′,5′-dideoxy adenosine.
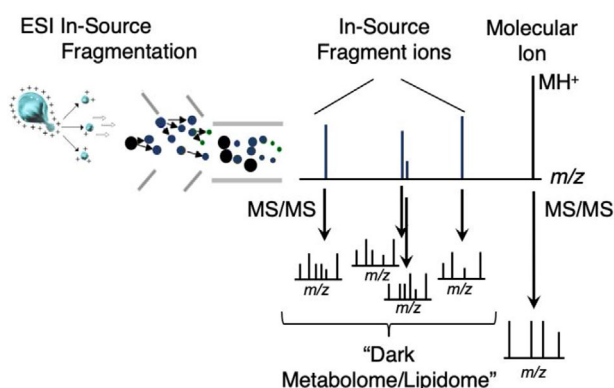
FIGURE 4 | ESI-ISF creates an overrepresentation of LC/MS peaks that are subjected to MS/MS analysis. This MS/MS data generated from ISF fragment ions, and the inability to identify them, appears to be largely responsible for the "Dark Metabolome/Lipidome".

Overall, these comparative examples between METLIN MS/MS (0 eV) data and ISF provide another level of evidence that the peaks observed in LC/MS experiments are predominantly associated with ISF. Figure 4 also illustrates the conceptual reasoning behind this logic, where MS/MS data are generated on all the unfiltered observable LC/MS peaks. Given the prevalence of ISF, most of the MS/MS data do not represent real molecules but instead fragment ions. This would explain why so many peaks are not identifiable in current tandem mass spectrometry databases.

Author Contributions

Winnie Uritboonthai: Data curation, formal analysis. **Linh Hoang:** Data acquisition. **Aries Aisporna:** Software. **Martin Giera:** Writing–review & editing. **Gary Siuzdak:** Experimental design, formal analysis, writing.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. M. Giera, A. Aisporna, W. Uritboonthai, and G. Siuzdak, "The Hidden Impact of in-source Fragmentation in Metabolic and Chemical Mass Spectrometry Data Interpretation," *Nature Metabolism* 6 (2024): 1647–1648, https://doi.org/10.1038/s42255-024-01076-x.

2. B. Y. L. Peisl, E. L. Schymanski, and P. Wilmes, "Dark Matter in Host-microbiome Metabolomics: Tackling the Unknowns–A Review," *Analytica Chimica Acta* 1037 (2018): 13–27, https://doi.org/10.1016/j.aca.2017.12.034.

While the METLIN study provides a large statistical snapshot of the number of ISF peaks in a typical LC/MS experiment, it lacked example data directly comparing the similarity between ISF and MS/MS (0 eV) data. Here, we examined both types of data (METLIN MS/MS 0 eV and ISF) from 10 molecules (Figures 2 and 3). ISF data were acquired using both an Agilent QTOF (collision cell off) mass spectrometer and an Agilent TOF mass spectrometer. The data revealed a high level of consistency between METLIN MS/MS 0 eV and ISF produced fragment ions, although the intensities were generally higher for the ISF generated fragment ion peaks. These examples suggest that (1) the original comparison between ISF and MS/MS (0 eV) is valid, and (2) the higher intensities observed for ISF fragments indicate that ISF process is slightly more energetic than MS/MS (0 eV), at least with these two instrument platforms (Agilent QTOF and Agilent TOF).

3. R. R. da Silva, P. C. Dorrestein, and R. A. Quinn, "Illuminating the Dark Matter in Metabolomics," *Proceedings of the National Academy of Sciences* 112 (2015): 12549–12550, https://doi.org/10.1073/pnas.1516878112.

4. X. Domingo-Almenara, J. R. Montenegro-Burke, C. Guijas, and E. L. Majumder, "Autonomous METLIN-Guided In-Source Fragment Annotation for Untargeted Metabolomics," *Analytical Chemistry* 91 (2019): 3246–3253, https://doi.org/10.1021/acs.analchem.8b03126.

5. J. Xue, X. Domingo-Almenara, and C., e. a. Guijas, "Enhanced In-Source Fragmentation Annotation Enables Novel Data-independent Acquisition and Autonomous METLIN Molecular Identification," *Analytical Chemistry* 92 (2020): 6051–6059, https://doi.org/10.1021/acs.analchem.0c00409.

6. S. Bernardo-Bermejo, J. Xue, and L., e. a. Hoang, "Quantitative Multiple Fragment Monitoring With Enhanced in-source Fragmentation/Annotation Mass Spectrometry," *Nature Protocols* 18 (2023): 1296–1315, https://doi.org/10.1038/s41596-023-00803-0.

7. P. Amaral, S. D. Sala, F. M. De La Vega, et al., "The Status of the human Gene Catalogue," *Nature* 622 (2023): 41–47, https://doi.org/10.1038/s41586-023-06490-x.

8. A. Varabyou, M. J. Sommer, B. Erdogdu, et al., "CHESS 3: An Improved, Comprehensive Catalog of Human Genes and Transcripts Based on Large-Scale Expression Data, Phylogenetic Analysis, and Protein Structure," *Genome Biology* 24, (2023) 249, https://doi.org/10.1186/s13059-023-03088-4.

9. J. M. Crow, "Canada's Scientists Are Elucidating the Dark Metabolome," *Nature* 599 (2021): S14–15, https://doi.org/10.1038/d41586-021-03062-9.

10. F. Crick, "Central Dogma of Molecular Biology," *Nature* 227 (1970): 561–563, https://doi.org/10.1038/227561a0.

11. Y. F. Xu, W. Lu, and J. D. Rabinowitz, "Avoiding Misannotation of in-source Fragmentation Products as Cellular Metabolites in Liquid Chromatography-mass Spectrometry-based Metabolomics," *Analytical Chemistry* 87 (2015): 2273–2281, https://doi.org/10.1021/ac504118y.

12. L. Chen, Lin Chen, and Hong Pan, "Widespread Occurrence of in-source Fragmentation in the Analysis of Natural Compounds by Liquid Chromatography–electrospray Ionization Mass Spectrometry," *Rapid Communications in Mass Spectrometry* 37 (2023): e9519, https://doi.org/10.1002/rcm.9519.

13. S. Bernardo-Bermejo, "Quantitative Multiple Fragment Monitoring With Enhanced In-Source Fragmentation/Annotation Mass Spectrometry," *Nature Protocols* 18 (2023): 1296–1315, https://doi.org/10.1038/s41596-023-00803-0.

14. C. Hoang, "Tandem Mass Spectrometry Across Platforms," *Analytical Chemistry* 96 (2024): 5478–5488, https://doi.org/10.1021/acs.analchem.3c05576.