Studying Charge Migration Fragmentation of Sodiated Precursor Ions in Collision-Induced Dissociation at the Library Scale

Marcus Ludwig[1], Corey D. Broeckling[2], Pieter C. Dorrestein[3], Kai Dührkop[1], Emma L. Schymanski[4], Sebastian Böcker[1,] Louis-Félix Nothias[3,*]

1) Chair for Bioinformatics, Friedrich Schiller University, 07743 Jena, Germany

2) Analytical Resources Core: Bioanalysis and Omics Center, Colorado State University, Fort Collins, CO 80523, USA

3) Collaborative Mass Spectrometry Innovation Center, and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, 92093, California, USA

4) Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg

THIS IS A REPRINT OF:

M. Ludwig, C. D. Broeckling, P. C. Dorrestein, K. Dührkop, E. L. Schymanski, S. Böcker, and L.-F. Nothias. Studying charge migration fragmentation of sodiated precursor ions in collision-induced dissociation at the library scale.

J Am Soc Mass Spectrom, 32(1):180–186, 2021.

https://doi.org/10.1021/jasms.0c00240

KEYWORDS: Tandem Mass Spectrometry, Collision Induced Dissociation, Fragmentation Mechanism, Library-Wide Fragmentation Mining.

Abstract: Interpretation of fragmentation mass spectra depends on our knowledge of collision-induced dissociation mechanisms. Computational methods for the annotation of fragmentation mechanisms operate within the boundaries of recognized fragmentation pathways. The prevalence of charge migration fragmentation (CMF) in sodiated ion fragmentation spectra, which produces nonsodiated fragment ions, is unknown. Here, we investigated the extent of CMF in the fragmentation spectra of sodiated precursors by mining the NIST17 spectral library using a diagnostic mass difference. Our results showed that a substantial amount of fragment ions in sodiated precursor spectra are derived from CMF, indicating that this fragmentation mechanism should be commonly considered for compound annotation.

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is one of the most common analytical platforms used for metabolomics and other small molecule analyses.[1] Although initial annotation of LC-MS/MS data can be performed at the MS1 level (accurate mass, isotope pattern, adduct identification, retention time), more confidence in the annotation can be obtained when supplemented by fragmentation information from e.g. tandem mass spectrometry (MS/MS) data collected in non-targeted mode.[2] MS/MS spectra are usually searched against spectral libraries, but typically only a few percent can be annotated.[3,4] The introduction of community-curated spectral libraries such as MassBank[5] and GNPS[6] facilitates the deposition of fragmentation spectra of annotated/identified compounds. Nevertheless, there is only a limited number of reference fragmentation spectra available for biomolecules.[7]

Computational methods that search in molecular structure databases have been developed to overcome the limitations of spectral libraries.[8] Combinatorial fragmentation methods such as MetFrag[9], MAGMa[10], DEREPLICATOR[11], and MS-Finder[12] attempt to explain fragment ions with substructures of a candidate molecule structure. SIRIUS[13] annotates an unknown query spectrum using fragmentation trees prior to predicting a molecular fingerprint with CSI:FingerID.[14] CFM-ID predicts hypothetical fragmentation spectra from molecular structures and compares these to the query spectrum to find the best molecular structure match.[15] All these approaches base their computations on certain assumptions about fragmentation processes, and incomplete or overly simplistic assumptions may reduce accuracy and performance dramatically. Thus, improving the general understanding of fragmentation patterns will help to improve computational annotation methods.

Fragmentation reactions in electrospray ionization (ESI) mass spectrometry are generally classified either as charge retention fragmentation (CRF) or as charge migration fragmentation (CMF).[16] The existence of CMF in the fragmentation of sodiated precursor ions ($[M+Na]^+$) is common knowledge for mass spectrometrists[16], but is often overlooked because it produces low intensity fragment ions. Recent studies on diterpene esters using multiple stage mass spectrometry[17–20] reported the presence of nonsodiated fragment ions indicative of CMF in the fragmentation spectra of $[M+Na]^+$. In ESI, these plant metabolites often form $[M+Na]^+$ adducts and produce rich fragmentation spectra that are dominated by sodiated fragment ions following hydrogen rearrangement by CRF. Nevertheless, the nonsodiated fragment ions from the CMF were found to be essential for the annotation of the diterpene ester

backbone.[20,21] The comparatively low relative and the importance of these CMF ions for the interpretation of these fragmentation spectra raises the question of how frequently the CMF pathway occurs in for sodiated precursor ions.

In this article, we studied the fragmentation patterns of CRF and CMF mechanisms in the CID spectra of $[M+Na]^+$. It was demonstrated that CRF and CMF occur jointly in many fragmentation spectra of $[M+Na]^+$ ions and that this joint occurrence manifests itself by a mass difference corresponding to one sodium minus one hydrogen (Na-H, 21.9819 Da). Searching for this specific mass difference in the NIST17 spectral library allowed us to estimate the frequency of the CRF/CMF mechanism in the fragmentation of $[M+Na]^+$, and the chemical properties of molecules favoring these fragmentation pathways.

MATERIALS AND METHODS

Fragmentation pathway of a representative diterpene ester. The fragmentation spectrum of a representative diterpene ester, 12***β***-O-[deca-2Z,4E,6E-trienoyl]13-isobutyroyloxy 4***β***-deoxyphorbol, was obtained from the GNPS library (CCMSLIB00000840551) and used to annotate its fragmentation pathway. This compound was isolated from *Euphorbia dendroides*[19] and its structure elucidated by the interpretation of the 1D and 2D NMR spectroscopy data. The compound was further and analyzed by nontargeted LC-MS/MS on an LTQ-XL Orbitrap. The extract and fractions were also analysedand the data deposited on MassIVE (MSV000080502).[17] Other isolated diterpene esters with the same fragmentation behavior were characterized by LC-MS/MS[20,22,23] and can be found on the GNPS library, including phorboids and jatrophane diterpenoids.

Preparation and annotation of the NIST17 MS/MS Spectral Library. The NIST MS/MS 2017 (NIST17) spectral library was mined to explore the extent of the CMF in sodiated ion CID fragmentation.[24,25] The NIST17 library was exported as .MSP format using the *LIB2NIST* converter software (574,826 spectra). In the NIST17, molecular structures are labeled with their InChIKeys and chemical name; spectra are normalized to a base peak intensity of 999. Low resolution spectra were excluded. Fragmentation spectra were combined into a *merged spectrum* if these had the same precursor *m*/*z*, InChIKey and consecutive NIST IDs; these spectra correspond to different collision energies acquired for one compound on a specific analytical platform. Fragment ions were merged if their mass difference was lower than 5 ppm or 0.001 Da; intensities of merged fragment ions were summed, and the new *m*/*z* calculated as the intensity-weighted average. After the merging, only fragment ions

above 1 % intensity of the base peak were retained and intensities were normalized to sum to one. Pairs of sodiated and protonated spectra of the same molecular structure were established by matching InChIKeys. This resulted in 1,803 pairs. Merging fragment ions with 10 ppm and 0.002 Da has also been tested, but the average number of peaks per merged spectrum does not change significantly (34.70 peaks with 5 ppm and 34.60 with 10 ppm).

The classification of structures using the ChemOnt chemical ontology was performed by querying ClassyFire[26] with the InChIKeys, resulting in 1,735 annotated spectral pairs. Additionally, SMILES were retrieved for the compounds by searching PubChem with the InChIKeys (25 Nov 2019): a SMILES string was accepted if PubChem contained a single unambiguous SMILES for the compounds' InChIKey plus chemical name, resulting in 1,149 spectral pairs with SMILES annotations. Based on the SMILES annotations, functional groups were counted using ChemmineOB (version 2.30.2)[27] package in R (version 3.4.4).[28]

Analysis of the NIST17 spectral library. Fragment ions in the spectra from NIST17 were attributed to the CMF or CRF pathway by 1) matching fragment ions between the fragmentation spectra of protonated and sodiated precursor ion or 2) searching for a characteristic 21.9819 Da *mass difference* between fragment ions in the fragmentation spectrum of the $[M+Na]^+$ ion. A maximum error of 5 ppm and 0.001 Da was considered to match fragment ions or mass differences. In the following, only the exact *m*/*z* values are reported but mass errors are considered for all *m*/*z* comparisons. For the fragmentation spectrum of the $[M+Na]^+$ ion, it was assumed that a fragment ion originated from the CMF pathway if it matched another fragment ion in the $[M+H]^+$ spectrum; these are called "*shared fragment ions"* below. The other way around, a fragment ion was assumed to originate from the CRF pathway if another fragment ion had a mass difference of 21.9819 Da between the paired spectra, corresponding to the mass of one sodium minus one hydrogen (Na-H). This characteristic Na-H *mass difference* was searched directly within the $[M+Na]^+$ merged spectrum or searched between the merged spectra of $[M+H]^+$ and $[M+Na]^+$ ions. One fragment ion was counted per matched fragment pair. Hence, when searching Δ(Na-H) within a spectrum (not between pairs of spectra), the relative number of matched fragments cannot exceed 50%; for intensity statistics of these fragment pairs within spectra, the intensity of the lower fragment (the one presumably originating from the CMF pathway) was used.

To evaluate the influence of different collision energies, fragmentation spectra were categorized into low (≤15 V), medium (>15 and ≤30 V) and high (>30 V) collision energies prior to merging spectra of the same compound. Only

compounds with at least one $[M+Na]^+$ spectrum in each category were kept; merged spectra were created individually for each category; $[M+H]^+$ spectra were not partitioned. This resulted in 703 merged spectrum pairs.

RESULTS AND DISCUSSION

First, the CMF mechanism that leads to nonsodiated fragment ion(s) was examined in more detail. To do so, the fragmentation spectrum of the $[M+Na]^+$ ion formed by a representative diterpene ester was used to interpret the fragmentation mechanism pathways occurring under CID (Figure 1, and Supplementary Figure S1). The results showed that the four most intense fragment ions are produced by a CRF mechanism, producing sodiated fragment ions involving intramolecular proton rearrangement(s).

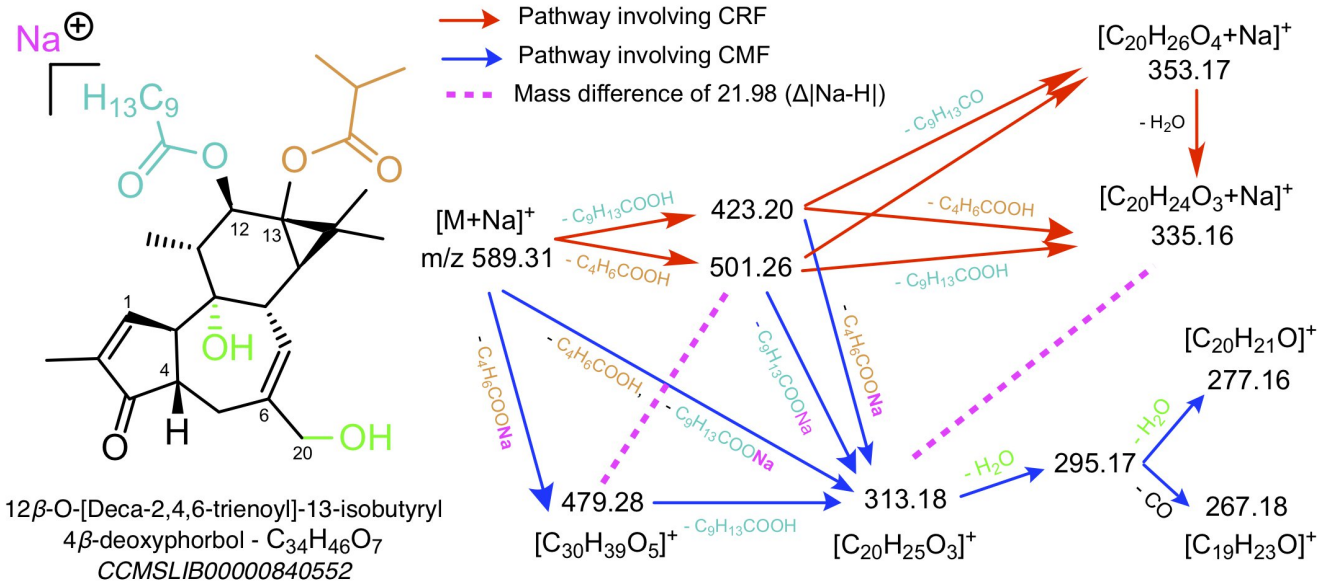

Figure 1. Fragmentation pathways occurring in collision-induced dissociation for the sodiaded precursor ion of the 12β-O-[deca-2Z,4E,6E-trienoyl] 13-isobutyroyloxy 4β-deoxyphorbol under a 30V normalized collision energy (See the spectrum in Supplementary Figure S1).

These CRF fragment ions (*m/z* 501.2642, 423.2148, 353.1735, and 335.1596) account for 88.5 % of the total ion count (TIC) in the studied spectrum. The CMF pathway produced low intensity nonsodiated fragment ions. The five most intense nonsodiated fragment ions produced by CMF account for only 5.4% of the TIC (*m/z* 479.2771, 313.1791, 295.1686, 277.1593, and 267.1759), which suggests that this CMF pathway is thermodynamically less

accessible than the CRF pathway for this sodiated ion. Nevertheless, the fragment ions produced by CMF were found to be critical for the determination of the diterpene backbone.[20,21] The present analysis here indicates that a 21.9819 Da mass difference is characteristic of a related pair of fragment ions produced by CRF (sodiated fragment ion) and CMF (nonsodiated fragment ion), such as *m/z* 501.2641 and 479.2771, or *m/z* 335.1597 and 313.1791. This same mass difference was also frequently observed in the majority of $[M+Na]^+$ spectra of other diterpene esters deposited on the GNPS library (Figure S1-S2), supporting the premise that the approach is capable to detect CMF fragment ion(s).

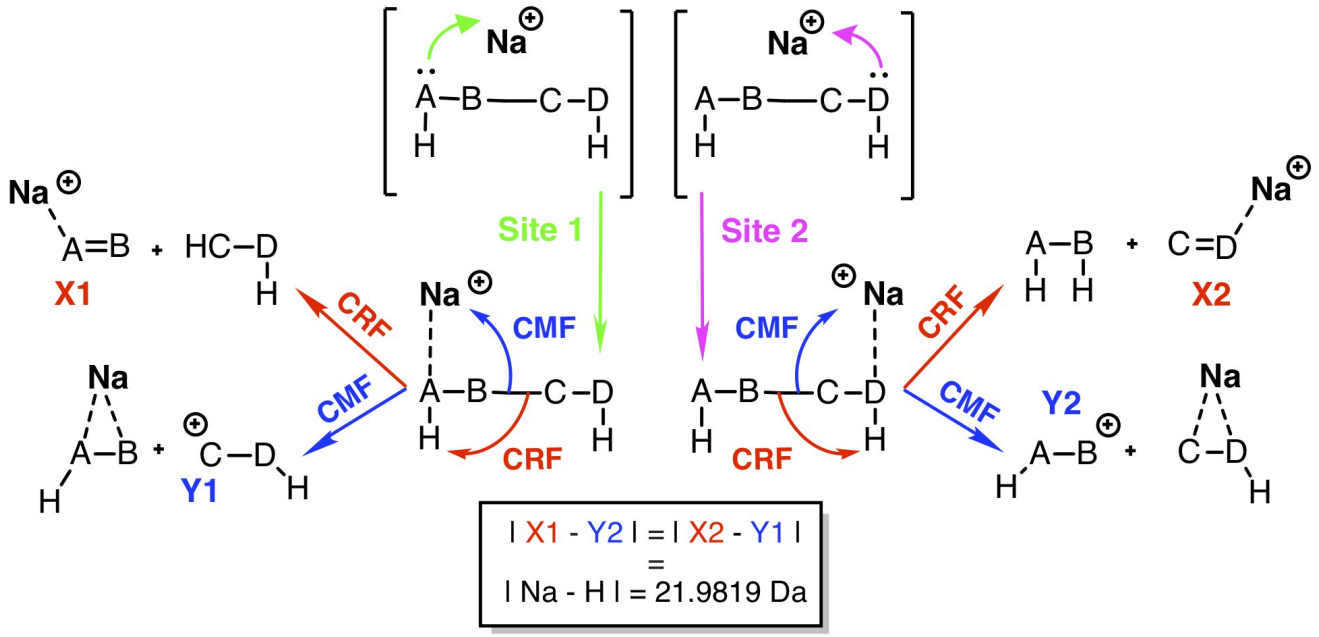

Figure 2. Representation of the fragmentation pathways occurring in collision-induced dissociation of sodiated precursor in ESI when two ionisation sites are possible (1 and 2). Charge retention fragmentation (CRF) via remote hydrogen rearrangement is the most frequent mechanism and produces sodiated fragment ions. Charge migration fragmentation (CMF) produces nonsodiated fragment ions following a neutral elimination of the sodiated fragment. A difference of 21.9819 Da in a $[M+Na]^+$ fragmentation spectrum is characteristic of the simultaneous occurrence of the two fragmentation pathways occurring on different ionisation sites.

Figure 2 describes the simultaneous occurrence of the CRF and CMF pathways in $[M+Na]^+$ fragmentation spectra. The CRF pathway involves an intramolecular proton rearrangement resulting in sodiated fragment ions

and the elimination of a neutral fragment of the molecule. This is the best-known fragmentation mechanism for sodiated ions.[16] The CMF is caused by an intramolecular cleavage involving an intramolecular transfer of electrons from the molecule to the sodiated fragment, leading to the production of nonsodiated fragment ion(s), after the neutral loss of a sodiated fragment ion as the sodium cation ionically bonds an anionic molecular fragment.

To better understand the prevalence of CMF in the fragmentation spectra of small molecules, MS/MS reference spectra from the NIST17 library were investigated. Whereas fragment ions cannot be identified directly as products of CMF or CRF without performing MS^n experiments, the corresponding fragmentation pathway can be inferred from the observation of mass differences between fragment ions. To do so, the fragmentation spectra of molecules with both sodiated and protonated ion spectra were examined (Figure 3a).

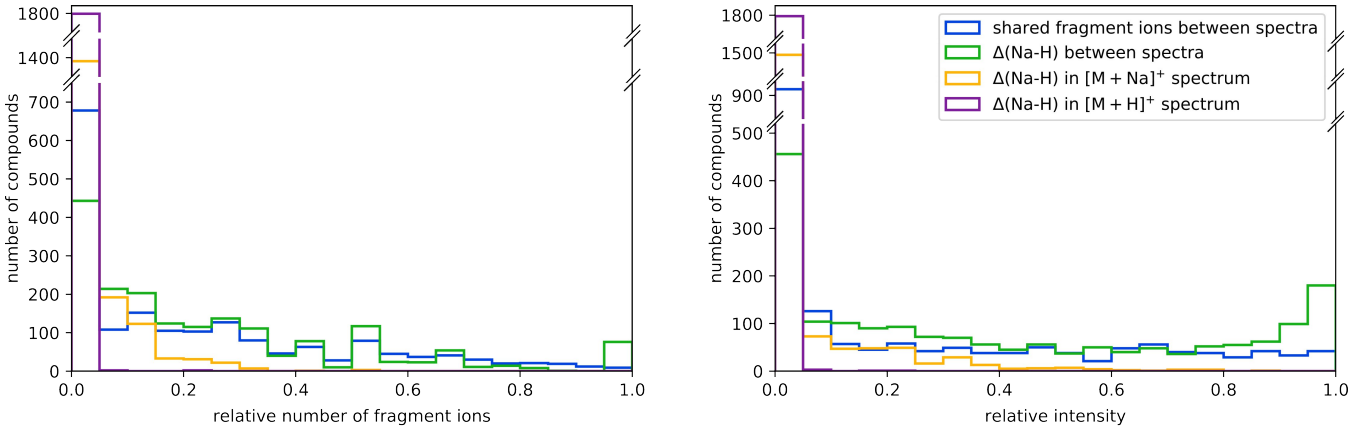

Figure 3. Histogram of (a) the relative number of fragment ions in the spectrum that can be matched between pairs of $[M+H]^+$ and $[M+Na]^+$ fragmentation spectra and (b) their summed relative intensity to the total ion current: shared fragment ions between spectra (blue) with the same m/z value indicate the occurrence of CMF fragmentation in $[M+Na]^+$ spectra, while the Δ(Na-H) difference between fragment ions for the spectral pair (green) is indicative of sodiated fragment ions. In addition, the frequency of the Δ(Na-H) mass difference in $[M+Na]^+$ spectrum was investigated (gold). The same Δ(Na-H) difference was also searched within the $[M+H]^+$ spectrum (violet); such matches represent false positives. Note the broken y-axis.

Shared fragment ions that are observed in both $[M+H]^+$ and $[M+Na]^+$ ion fragmentation spectra (0 Da difference) are nonsodiated. In the $[M+Na]^+$ fragmentation spectrum, these shared fragment ions must originate from the CMF pathway. On the contrary, if the fragmentation spectra of a $[M+H]^+$ and $[M+Na]^+$ spectral pair contain a mass

difference of 21.9819 Da - corresponding to the mass of sodium minus hydrogen (Na-H) - it is indicative of sodiated fragment ions originating from the CRF pathway in the $[M+Na]^+$ spectrum. Moreover, the characteristic Na-H difference was searched directly within the sodiated ion spectra to find related pairs of nonsodiated and sodiated fragment ions. Our results showed that 24.35% of fragment ions in the $[M+Na]^+$ spectra are matched by a Δ(Na-H) mass difference to a counterpart fragment ion in $[M+H]^+$ spectra. Remarkably, it could be observed that 21.76% of fragment ions in the $[M+Na]^+$ are actually shared between the $[M+H]^+$ and $[M+Na]^+$ fragmentation spectra. Only 0.01% of all fragment ions in $[M+Na]^+$ spectra match ambiguously. In addition, results showed that fragment ions differing by Δ(Na-H) within the fragmentation spectra of $[M+Na]^+$ precursor ions are less frequent, showing that CMF and CRF pathways are not always observed simultaneously in $[M+Na]^+$ spectrum. Investigating the effect of collision energy showed that a higher energy slightly increases the frequency of observing fragment ions from the CMF pathway (Figure S3) in these sodiated species. The relative proportion of annotated fragment ions from the CRF and CMF pathway is presented in Figures S4 and S5. To exclude that two fragment ions have a Δ(Na-H) mass difference just by chance, the same difference was searched within the protonated ion spectra, which should not include such fragment pairs. For less than 1.7% of the protonated ion spectra this resulted in more than 1% of matched fragment ions, showing that fragment ions are rarely matching by chance.

The summed relative intensity of fragment ions for each pathway was examined (Figure 3b). The majority of $[M+Na]^+$ spectra were dominated by sodiated fragment ions from CRF, while few contained primarily nonsodiated ions. The distribution of the relative intensities for the individual fragment ions shows that the CMF pathway in the $[M+Na]^+$ spectra tends to produce lower intensity fragment ions than the CRF pathway (Figure S6). In addition, we found that Na-H mass difference was the 11th most frequent (0.44%) in the fragmentation spectra of $[M+Na]^+$ ions (Table S1).

Next, to investigate if some chemical properties were associated with CMF in $[M+Na]^+$, frequent mass differences between $[M+H]^+$ and $[M+Na]^+$ fragmentation spectra were searched. In total, one million mass differences between fragment ions of spectral pairs were randomly sampled. Sampled values were rounded to 3 decimal places and counted to select the most frequent mass differences (Table S2). Results showed the most frequent mass differences were 21.982 Da corresponding to Δ(Na-H) (1.27%), and 0.000 Da (0.91%), along with mass differences indicative of Δ(Na-H) plus loss of water (39.992 Da, 0.51%) and Δ(Na-H) plus a loss of a carbonyl (49.977, 0.23%).

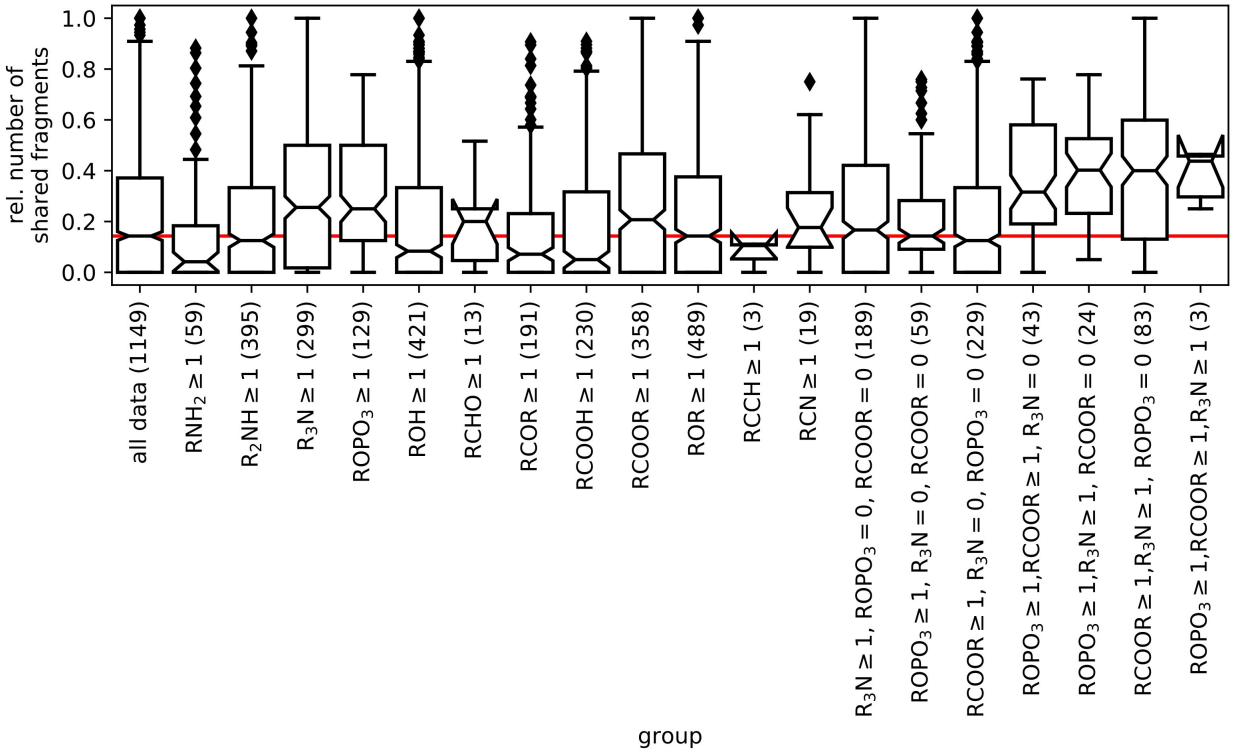

Figure 4. Ratio of shared fragment ions between $[M+H]^+$ and $[M+Na]^+$ spectra of the same compound depending on the existence of various functional groups. Shared fragment ions are used as an approximation for the ratio of nonsodiated fragment ions in $[M+Na]^+$spectra. The boxes indicate the 25th and 75th percentile, the median and its confidence interval. The red line marks the median ratio of shared fragment ions for all available data. For RCOOR, $ROPO_3$ and R_3N groups further combinations are displayed. The numbers in brackets indicate the sample size.

The relationship between chemical functions and CMF in sodiated ions was studied by looking at the frequency of functional groups for whether their presence increases the probability of observing a CMF fragment ions (Figure 4). Results showed that fragment ions indicative of CMF were relatively ubiquitous in compounds containing at least one heteroatom. Oppositely, CMF fragment ions were not observed in alkanes. The combination of at least two of the functional groups RCOOR, $ROPO_3$ and R_3N favored nonsodiated fragment ions. Thus, two situations could be favoring CMF: (1) acidic functional groups presumably acting as sodium cation counterions or (2) basic moieties on the charged fragment, which would favor intramolecular cation formation.

Additionally, it was investigated whether some chemical classes are more prone to CMF. Therefore, compounds were classified using the ChemOnt chemical ontology.[26] The distribution of compound superclasses is summarized in Figure 5 and showed the proportion of $[M+Na]^+$ spectra enriched with nonsodiated fragments (664 compounds with at least 20% relative number of nonsodiated fragment ions) per chemical superclass. For robustness, the data

was limited to spectra with at least 5 fragment ions. Fragmentation spectra of sodium precursors in the NIST17 library encompass many ClassyFire superclasses. This diversity suggests that CMF is widespread in the fragmentation spectra of $[M+Na]^+$ ions. Compounds belonging to the superclasses "nucleosides, nucleotides, and analogues" and "organic nitrogen compounds" had a higher frequency of observable CMF derived fragment ions in their sodiated precursor ion fragmentation spectra. This agrees with the observations made above that functional groups $ROPO_3$ and R_3N favor CMF in fragmentation spectra of sodiated ions.

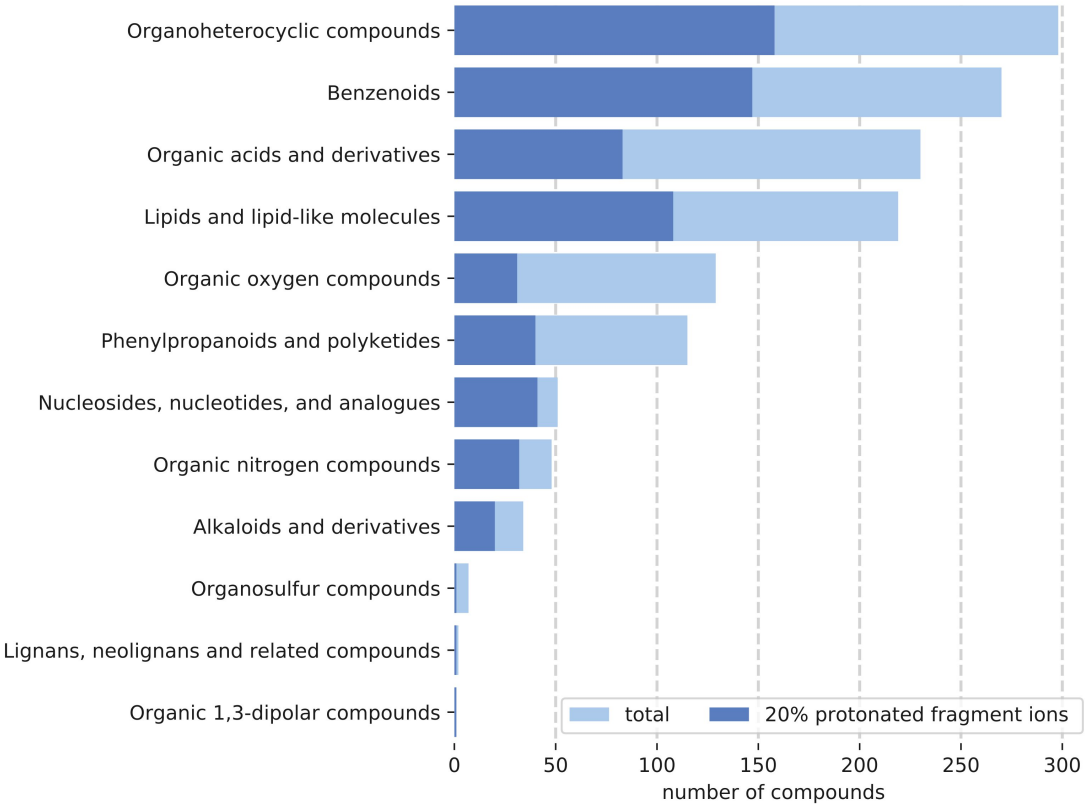

Figure 5. Distribution of the $[M+Na]^+$ spectra in the NIST17 library. Superclass membership in the ChemOnt chemical ontology for all compounds, and the proportion of spectra with more than 20% of nonsodiated fragments (estimated by searching shared fragments between $[M+H]^+$ and $[M+Na]^+$ spectral pairs). The occurrence of nonsodiated fragment ions in $[M+Na]^+$ spectra is widespread with respect to chemical superclass.

CONCLUSIONS

The presented results support the frequent occurrence of CMF in the fragmentation spectra of $[M+Na]^+$ ions across many chemical classes of small molecules. The scale of this phenomenon stresses that it should be considered during the design of computational methods for the annotation of $[M+Na]^+$ fragmentation spectra. Improving our

ability to detect and annotate the fragmentation spectra of $[M+Na]^+$ ions will facilitate the addition of new $[M+Na]^+$ fragmentation spectra into spectral libraries, that in turn will improve the efficiency of computational annotation methods in metabolomics. Although sodiated precursor ions represent ~25% of the adducts that can be identified in LC-MS-based metabolomics experiments,[29] these adducts only represent ~10 % of spectral libraries entries. In addition, because adduct annotation based on MS^1-only can be ambiguous, our results indicate that, when fragmentation spectra are available, the Na-H mass difference could be used to add confidence in identifying the ion as a sodium adduct. Nevertheless, it should be noted that the presented analysis, in particular on functional groups and compound classes, is to some extent biased by the composition of the NIST17 MS/MS library. From a methodological standpoint, this study illustrates that by leveraging spectral libraries, it is possible to contextualize experimental observation in collision-induced dissociation. We anticipate that this complementary strategy will gain in usefulness with the increased availability of reference fragmentation spectra in public libraries.

ASSOCIATED CONTENT

Supporting Information

A PDF file containing annotated fragmentation spectra, additional statistical analysis conducted in this study, and data from the structures from NIST17 that were analyzed. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

* Contact information for the author to whom correspondence should be addressed: lnothiasscaglia@health.ucsd.edu

Author Contributions

ML and LFN performed the analysis, with the help of all authors. ML, CB, ES, SB, and LFN wrote the manuscript.

ACKNOWLEDGMENT

LFN was supported by the U.S. National Institutes of Health (R01 GM107550), and the European Union's Horizon 2020 program (MSCA-GF, 704786). PCD was supported by the Gordon and Betty Moore Foundation (GBMF7622), the U.S. National Institutes of Health (P41 GM103484, R03 CA211211, R01 GM107550). ELS acknowledges funding from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006. SB, KD and ML acknowledge financial support by the Deutsche Forschungsgemeinschaft (BO 1910/20). CDB was supported by the U.S. National Institutes of Health (1U01CA235507-01) and acknowledges the CSU Office of the Vice President for Research.

REFERENCES

(1) Wishart, D. S. Emerging Applications of Metabolomics in Drug Discovery and Precision Medicine. *Nat. Rev. Drug Discov.* **2016**, *15* (7), 473–484.
(2) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098.
(3) da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A. Illuminating the Dark Matter in Metabolomics. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12549–12550.
(4) Oberacher, H.; Sasse, M.; Antignac, J.-P.; Guitton, Y.; Debrauwer, L.; Jamin, E. L.; Schulze, T.; Krauss, M.; Covaci, A.; Caballero-Casero, N.; Rousseau, K.; Damont, A.; Fenaille, F.; Lamoree, M.; Schymanski, E. L. A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries. *Environmental Sciences Europe* **2020**, *32* (1), 43.
(5) 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* (7), 703–714.
(6) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; P, C. A. B.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters,

K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837.

(7) Frainay, C.; Schymanski, E. L.; Neumann, S.; Merlet, B.; Salek, R. M.; Jourdan, F.; Yanes, O. Mind the Gap: Mapping Mass Spectral Databases in Genome-Scale Metabolic Networks Reveals Poorly Covered Areas. *Metabolites* **2018**, *8* (3). https://doi.org/10.3390/metabo8030051.

(8) Hufsky, F.; Scheubert, K.; Böcker, S. New Kids on the Block: Novel Informatics Methods for Natural Product Discovery. *Nat. Prod. Rep.* **2014**, *31* (6), 807.

(9) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminform.* **2016**, *8* (1), 1–16.

(10) Ridder, L.; van der Hooft, J. J. J.; Verhoeven, S. Automatic Compound Annotation from Mass Spectrometry Data Using MAGMa. *Mass Spectrom.* **2014**, *3* (Spec Iss 2), S0033.

(11) Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L.-F.; Ninomiya, A.; Takada, K.; Dorrestein, P. C.; Pevzner, P. A. Dereplication of Peptidic Natural Products through Database Search of Mass Spectra. *Nat. Chem. Biol.* **octobre 31, 2016**, No. 1, 30–37.

(12) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* **2015**, *12* (6), 523–526.

(13) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16* (4), 299–302.

(14) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12580–12585.

(15) Allen, F.; Greiner, R.; Wishart, D. Competitive Fragmentation Modeling of ESI-MS/MS Spectra for Putative Metabolite Identification. *Metabolomics* **2015**, *11* (1), 98–110.

(16) Demarque, D. P.; Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P. Fragmentation Reactions Using Electrospray Ionization Mass Spectrometry: An Important Tool for the Structural Elucidation and Characterization of Synthetic and Natural Products. *Nat. Prod. Rep.* **2016**. https://doi.org/10.1039/C5NP00073D.

(17) Esposito, M.; Nothias, L.-F.; Nedev, H.; Gallard, J.-F.; Leyssen, P.; Retailleau, P.; Costa, J.; Roussi, F.; Iorga, B. I.; Paolini, J.; Litaudon, M. Euphorbia Dendroides Latex as a Source of Jatrophane Esters: Isolation, Structural Analysis, Conformational Study, and Anti-CHIKV Activity. *J. Nat. Prod.* **2016**, *79* (11), 2873–2882.

(18) Esposito, M.; Nothias, L.-F.; Retailleau, P.; Costa, J.; Roussi, F.; Neyts, J.; Leyssen, P.; Touboul, D.; Litaudon, M.; Paolini, J. Isolation of Premyrsinane, Myrsinane, and Tigliane Diterpenoids from Euphorbia Pithyusa Using a Chikungunya Virus Cell-Based Assay and Analogue Annotation by Molecular Networking. *J. Nat. Prod.* **2017**, *80* (7), 2051–2059.

(19) Nothias, L.-F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C. Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J. Nat. Prod.* **2018**. https://doi.org/10.1021/acs.jnatprod.7b00737.

(20) Nothias-Scaglia, L.-F.; Schmitz-Afonso, I.; Renucci, F.; Roussi, F.; Touboul, D.; Costa, J.; Litaudon, M.; Paolini, J. Insights on Profiling of Phorbol, Deoxyphorbol, Ingenol and Jatrophane Diterpene Esters by High Performance Liquid Chromatography Coupled to Multiple Stage Mass Spectrometry. *J. Chromatogr. A* **2015**, *1422*, 128–139.

(21) Vogg, G.; Achatz, S.; Kettrup, A.; Sandermann, H. Fast, Sensitive and Selective Liquid Chromatographic–tandem Mass Spectrometric Determination of Tumor-Promoting Diterpene Esters. *J. Chromatogr. A* **1999**, *855* (2), 563–573.

(22) Nothias, L.-F.; Boutet-Mercey, S.; Cachet, X.; De La Torre, E.; Laboureur, L.; Gallard, J.-F.; Retailleau, P.; Brunelle, A.; Dorrestein, P. C.; Costa, J.; Bedoya, L. M.; Roussi, F.; Leyssen, P.; Alcami, J.; Paolini, J.; Litaudon, M.; Touboul, D. Environmentally Friendly Procedure Based on Supercritical Fluid Chromatography and Tandem Mass Spectrometry Molecular Networking for the Discovery of Potent Antiviral Compounds from Euphorbia Semiperfoliata. *J. Nat. Prod.* **2017**, *80* (10), 2620–2629.

(23) Nothias-Esposito, M.; Nothias, L. F.; Da Silva, R. R.; Retailleau, P.; Zhang, Z.; Leyssen, P.; Roussi, F.; Touboul, D.; Paolini, J.; Dorrestein, P. C.; Litaudon, M. Investigation of Premyrsinane and Myrsinane Esters in Euphorbia Cupanii and Euphobia Pithyusa with MS2LDA and Combinatorial Molecular Network Annotation Propagation. *J. Nat. Prod.* **2019**, *82* (6), 1459–1470.
(24) Yang, X.; Neta, P.; Stein, S. E. Extending a Tandem Mass Spectral Library to Include MS2 Spectra of Fragment Ions Produced In-Source and MSn Spectra. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2280–2287.
(25) National Institute of Standards and Technology, US Secretary of Commerce, Gaithersburg, Maryland, USA. NIST/EPA/NIH Mass Spectral Library v2017 https://www.nist.gov/srd/nist-standard-reference-database-1a-v17 (accessed Sep 4, 2018).
(26) Feunang, Y. D.; 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* (1), 61.
(27) Cao, Y.; Charisi, A.; Cheng, L.-C.; Jiang, T.; Girke, T. ChemmineR: A Compound Mining Framework for R. *Bioinformatics* **2008**, *24* (15), 1733–1734.
(28) Team, R. C.; Others. R: A Language and Environment for Statistical Computing. **2013**.
(29) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; Korf, A.; Pluskal, T.; Kameník, Z.; Jarmusch, A. K.; Caraballo-Rodríguez, A. M.; Weldon, K.; Nothias-Esposito, M.; Aksenov, A. A.; Bauermeister, A.; Orio, A. A.; Grundmann, C. O.; Vargas, F.; Koester, I.; Gauglitz, J. M.; Gentry, E. C.; Hövelmann, Y.; Kalinina, S. A.; Pendergraft, M. A.; Panitchpakdi, M. W.; Tehan, R.; Le Gouellec, A.; Aleti, G.; Russo, H. M.; Arndt, B.; Hübner, F.; Hayen, H.; Zhi, H.; Raffatellu, M.; Prather, K. A.; Aluwihare, L. I.; Böcker, S.; McPhail, K. L.; Humpf, H.-U.; Karst, U.; Dorrestein, P. C. Ion Identity Molecular Networking in the GNPS Environment. *bioRxiv*, 2020, 2020.05.11.088948. https://doi.org/10.1101/2020.05.11.088948.