

# Enhanced Electrospray In-source Fragmentation for Higher Sensitivity Data Independent Acquisition and Autonomous METLIN Molecular Identification

Jingchuan Xue<sup>†,#</sup>, Xavier Domingo-Almenara<sup>‡,#</sup>, Carlos Guijas<sup>†</sup>, Amelia Palermo<sup>†</sup>, Markus Rinschen<sup>†</sup>, H. Paul Benton<sup>†</sup>, and Gary Siuzdak<sup>†,‡,\*</sup>

<sup>†</sup> Scripps Center for Metabolomics and <sup>‡</sup>Department of Molecular and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, United States

<sup>‡</sup> Omic Sciences Unit, EURECAT – Technology Centre of Catalonia, Reus, Catalonia, Spain

<sup>#</sup>authors contributed equally

\*Corresponding author

Gary Siuzdak, PhD

Scripps Center for Metabolomics

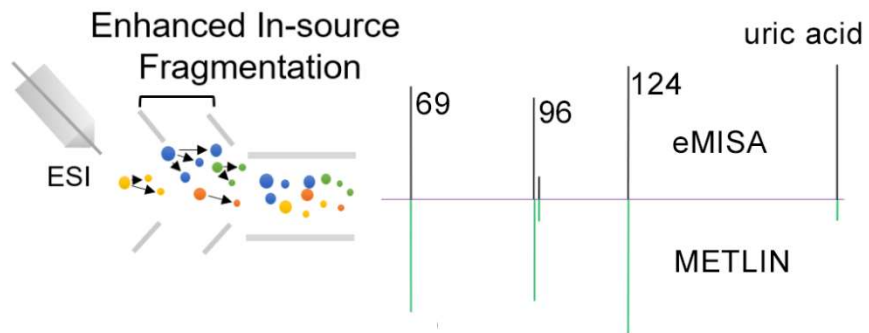
The Scripps Research Institute, La Jolla, CA 92037

Tel.: 858 784 9415

Email: [siuzdak@scripps.edu](mailto:siuzdak@scripps.edu)

For submission to: Analytical Chemistry

# TOC



**ABSTRACT:** Electrospray ionization (ESI) in-source fragmentation (ISF) has traditionally been minimized to promote precursor molecular ion formation, and therefore its value in molecular identification underappreciated. Recently a METLIN-guided in-source annotation (MISA) algorithm was introduced to increase confidence in putative identifications by using ubiquitous in-source fragments. However, MISA is limited by ESI sources that are generally designed to minimize ISF. In this study, enhanced ISF with MISA (eMISA) was created by tuning the ISF conditions to generate in-source fragmentation patterns comparable with higher energy fragments generated at higher collision energies as deposited in the METLIN MS/MS library, without compromising the intensity of precursor ions (median loss  $\leq 10\%$  in both positive and negative ionization modes). The analysis of 50 molecules was used to validate the approach in comparison to MS/MS spectra produced via data dependent acquisition (DDA) and data independent acquisition mode (DIA) with quadrupole time-of-flight mass spectrometry (QTOF-MS). Enhanced ISF as compared to QTOF DDA, enables for higher peak intensities for the precursor ions (median: 18 times at negative mode and 210 times at positive mode), with the eMISA fragmentation patterns consistent with METLIN for over 90% of the molecules with respect to fragment relative intensity and  $m/z$ . eMISA also provides higher peak intensity as opposed to QTOF DIA with a median increase of 20% at negative mode and 80% at positive mode for all precursor ions. Metabolite identification with eMISA was also successfully validated from the analysis of a metabolic extract from macrophages. An interesting side benefit of enhanced ISF is that it significantly improved the compound identification confidence with low resolution single quadrupole mass spectrometry-based untargeted LC/MS experiments. Overall, enhanced ISF allowed for eMISA to be used as a more sensitive alternative to other QTOF DIA and DDA approaches, and further, it enables the acquisition of ESI TOF and ESI single quadrupole mass spectrometry instrumentation spectra with higher sensitivity and improved molecular identification confidence.

## INTRODUCTION

Untargeted liquid chromatography electrospray ionization (ESI) mass spectrometry (LC-MS) experiments aim to capture the information for all the small molecules present in a sample.<sup>1,2</sup> Because of the high sensitivity, selectivity, and broad dynamic range, it has become the dominant platform in molecular analysis.<sup>3-5</sup> However, the annotation and identification of the thousands of features typically detected in LC-MS experiments remains a critical challenge.<sup>6-8</sup> It is well known that in addition to molecular ions, LC-MS<sup>1</sup> data also contains adducts, isotopes, multimers, in-source fragments, and contaminants, etc., which can significantly impact annotation.<sup>6,7,9</sup>

In-source fragmentation (ISF) is a naturally occurring phenomenon in atmospheric ion sources.<sup>10-13</sup> ESI is considered among the softest ionization technologies with the least ISF, however, even in the ESI source, unintentional ISF widely exists.<sup>7,14</sup> The fragmentation pattern generated by ISF in the ESI source is similar to the low-energy collision-induced dissociation (CID) MS/MS spectra.<sup>7</sup> For example, ISF of coumaric acid and phosphocholine are similar to the 0 and 10 eV MS/MS spectra reported in the METLIN database, respectively.<sup>7</sup> These lower molecular weight in-source fragment ions may lead to false positives in compound identification, especially when the in-source fragment ions correlate to the precursor ions of other molecules that elute at similar chromatographic retention times.<sup>14,15</sup> For example, pipecolic acid, an in-source fragment ion of lysine, elutes at the same retention time with lysine. ISF-induced mis-annotation has been widely observed in the literature. In another example, Xu *et al.* (2015) found that correct identification of 21 out of 130 routinely monitored water-soluble intracellular metabolites in the yeast metabolome are impacted by in-source fragment ions.<sup>14</sup> In-source cyclization of glutamine and glutamate to pyroglutamate have also been reported.<sup>16</sup> In a lipidomics study conducted in ESI negative ionization, 40 of the 100 most abundant masses corresponding to unique phospholipids in plasma metabolome were artifacts from ISF.<sup>15</sup>

Current efforts have been focused on minimizing or eliminating ISF to reduce the spectral complexity in untargeted LC-MS experiments by adjusting the in-source parameters.<sup>14,15</sup> However, completely removing the ISF is only possible by designing novel ionization techniques. Instead, the ESI in-source fragments present in the untargeted MS<sup>1</sup> data can be used for feature annotation.<sup>7,17-19</sup> In a previous study, our group developed an approach called METLIN-guided in-source annotation (MISA) specifically intended for ISF annotation using

experimental low-energy MS/MS spectra in the METLIN library.<sup>7</sup> MISA leverages the fact that in-source and low-energy MS/MS fragmentations are similar as they are generated by collision under electric fields.<sup>7</sup> In that sense, MISA consists in matching peaks observed in MS<sup>1</sup> data against low-energy MS/MS spectra from the METLIN database.<sup>7</sup> This allows a rapid metabolite annotation to facilitate preliminary data analysis and target MS/MS method design for metabolite identity confirmation.<sup>7</sup> However, under standard ion source conditions, ISF can only mimic low energy MS/MS spectra, typically from 0 to 10 eV.<sup>7</sup> Not all molecules undergo ISF at low energy, or they yield a small number of in-source fragments. This hampers the correct annotation of some molecules due to the low specificity of these fragments.

Currently, MS<sup>1</sup> metabolite profiling is normally conducted in the first step in a typical LC-MS based untargeted experiment, followed by MS/MS spectra generation in the collision cell. Data-dependent acquisition (DDA) is generally used to produce MS/MS spectra of a target precursor ion for compound identification by matching it to library spectra. However, due to stochastic nature of DDA, MS/MS spectra for some precursor ions cannot be generated, especially for those ions with low abundance. Alternatively, data-independent acquisition (DIA) techniques were designed to collect full MS/MS spectra on low abundance analytes, such as the SWATH acquisition in SCIEX instruments. However, it is difficult to link the fragmented precursor ion and its specific fragments in multiplexed MS/MS spectra using DIA approaches which makes processing DIA data more challenging.<sup>20</sup>

Here, we demonstrate that enhancing ISF enables acquiring the pseudo MS/MS spectra for a broad range of molecules comparable to those produced by DDA or DIA, thus demonstrating the utility of this method as an alternative to DIA methods without incurring losses of precursor ion intensity due to collision-induced dissociation within a collision cell and without the loss of the link between the precursor ion and its fragments as in DIA methods. In this study, we optimized the MISA workflow by enhancing ISF to generate metabolite in-source fragments that are comparable to the 20 eV MS/MS spectra in the METLIN library, as shown in Fig. 1. The performance of enhanced in-source fragmentation MISA (eMISA) was evaluated by the analysis of 50 endogenous metabolites using two indicators, the number of fragments that matched to the library (ratio score) and the relative intensity similarity between experimental and library mass spectra (match factor). Compared with the traditional MS/MS acquisition techniques, including both DDA and DIA, eMISA demonstrated comparable fragmentation

pattern, higher peak intensity for precursor ions, and better precursor ion sensitivity for the select molecules. Based on the MS<sup>1</sup> spectra generated with eMISA, we successfully identified metabolites in a macrophages extract, confirmed by the analysis of pure standards under the same analytical conditions. Further, the eMISA strategy was applied in the putative compound identification using a single quadrupole mass spectrometry-based untargeted data. Our results indicated that eMISA is a novel and much simplified DIA strategy for compound identification in LC-MS based untargeted experiments.

## EXPERIMENTAL SECTION

**Materials.** For the optimization of in-source fragmentation conditions, a mixture of 50 endogenous metabolites was prepared in water at 30  $\mu$ M. These metabolites were selected to represent a broad range of physicochemical properties and metabolic pathways, such as amino acids, lipids, citric acid cycle, nucleotides, and coenzyme. For the investigation of sensitivity, a mixture of four metabolites at each ionization mode (tyrosine, inosine, uridine monophosphate (UMP), and oxidized glutathione at negative mode; phenylalanine, glutamine, histidine, and oxidized glutathione at positive mode) was prepared at 9 concentrations ranging 6 orders of magnitude: 1 nM, 10 nM, 100 nM, 1  $\mu$ M, 10  $\mu$ M, 50  $\mu$ M, 100  $\mu$ M, 500  $\mu$ M, and 1 mM. The standards were purchased from Sigma-Aldrich (St. Louis, MO).

**Metabolite extraction.** Metabolites in the macrophage samples were extracted using the method described elsewhere with slight modifications.<sup>21</sup> In brief, cell samples were sonicated in ice for 15 mins after a shock-freezing in liquid nitrogen and subsequent thawing at room temperature. The operation was repeated for three times. Then the sample was incubated at – 20 °C overnight for protein precipitation, followed by centrifugation at 13, 000 rpm and 4 °C for 15 mins. The supernatant was dried in a vacuum concentrator and reconstituted with 50  $\mu$ L acetonitrile:water (v:v; 1:1). After sonication (10 mins, in ice) and centrifugation (13, 000 rpm, 4 °C, 15 mins), the supernatant was transferred to a LC-MS glass vial for instrumental injection.

**High-Resolution Mass Spectrometry Analysis.** The metabolite mixture was analyzed with a liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry (LC-ESI-QTOF-MS) (BRUKER impact II) in both positive and negative ESI modes. ZORBAX 300

SB-C18 column (0.5 × 150 mm, 5 μm, Agilent) and Luna® NH2 100 Å column (1 × 150 mm, 3 μm, Phenomenex) were used in the separation of metabolites in positive and negative modes, respectively. For the reverse phase analysis, metabolites were separated by gradient elution at a flow rate of 20 μL/min starting at 5% (v/v) B, held for 5 min, increased to 95% B within 45 min, held for 5 min, and reverted to 5% B at the 56<sup>th</sup> min, held for 5 min, with a total run time of 61 min. The mobile phases comprised water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B). For the HILIC analysis, metabolites were separated by gradient elution at a flow rate of 50 μL/min starting at 0% (v/v) A, held for 1 min, increased to 100% A within 47 min, held for 5 min, and reverted to 0% A at the 56<sup>th</sup> min, held for 5 min, with a total run time of 61 min. The mobile phases comprised water/acetonitrile (v/v; 95:5) containing 20 mM ammonium acetate and 40 mM ammonium hydroxide (A) and water/acetonitrile (v/v; 5:95) (B).

The original ESI source parameters were set as follows: dry temperature 180 °C, dry gas 7 L/min, nebulizer 20.3 psi, and capillary voltage 4500 V and – 4000 V for positive and negative modes, respectively. The transfer parameters were: funnel 1 RF 150 Vpp, funnel 2 RF 200 Vpp, isCID energy 0 eV, and hexapole RF 50 Vpp. The ion energy in the quadrupole was set to 4 eV and the collision energy applied in the collision cell was set at 7 eV. The instrument acquisition range was set at 50-1200 *m/z* and the MS acquisition rate was 2 spectra/s.

In the enhanced in-source fragmentation condition, transfer isCID energy increased to 30 eV at positive mode and 40 eV at negative mode, respectively, with all other parameters fixed.

For the data dependent acquisition (DDA) of MS/MS spectra, both the MS and MS/MS acquisition rates were set as 2 spectra/s to acquire over the *m/z* range 50-1200. The collision energy was fixed at 20 eV for all the molecules.

For the data independent acquisition (DIA) of MS/MS spectra, Broadband Collision Induced Dissociation (bbCID) mode was selected with low and high collision energy as 7 and 20 eV, respectively. isCID energy was set at 0 eV for both channel 1 (low collision energy) and 2 (high collision energy). The acquisition factors were set at 1 for both high and low collision energy. The low collision energy aims to provide full scan TOF MS spectra and the high collision energy is to generate the DIA MS/MS spectra for all the peaks without precursor isolation. The *m/z* range was set as 50-1200.

**Single Quadrupole Mass Spectrometry Analysis.** An Agilent InfinityLab Liquid Chromatography/Mass Selective Detector (LC-MSD) system was used in the single quadrupole MS based untargeted experiments. Poroshell 120 EC-C8 column (4.6 x 50 mm, 2.7  $\mu\text{m}$ , Agilent) was used in the analysis of metabolite mixture in the positive mode; Acquity BEH Amide column (1 x 100 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , Waters) was used in the negative mode. The gradient at positive mode was as follows: starting at 5% (v/v) B, held for 2 min, increased to 95% B within 12 min, held for 2.5 min, and reverted to 5% B at the 17<sup>th</sup> min, held for 3 min, with a total run time of 20 min; mobile phase A, water containing 0.1% formic acid (A); mobile phase B, acetonitrile containing 0.1% formic acid; mobile phase flow rate, 500  $\mu\text{L}/\text{min}$ . The gradient at negative mode was as follows: starting at 1% (v/v) A, held for 1 min, increased to 35% A within 13 min, then increased to 60% A at the 17<sup>th</sup> min, held for 1 min, and reverted to 1% A at the 19<sup>th</sup> min, held for 6 min, with a total run time of 25 min; mobile phase A, water:acetonitrile (v:v; 95:5) containing 5 mM ammonium acetate and 5 mM ammonium hydroxide; mobile phase B, water:acetonitrile (v:v; 5:95); mobile phase flow rate, 100  $\mu\text{L}/\text{min}$ .

The ESI source parameters were set as follows: at positive mode, drying gas flow 12 L/min, drying gas temperature 350  $^{\circ}\text{C}$ , nebulizer pressure 35 psig, and capillary voltage 4000 V; at negative mode, drying gas flow 8 L/min, drying gas temperature 325  $^{\circ}\text{C}$ , nebulizer pressure 30 psig, and capillary voltage 3500 V. Mass spectra was acquired over the m/z range of 50-1200 at two fragmentor voltages, 150 and 300 V, respectively.

**Data Analysis.** LC-MS and MS/MS datasets generated under different conditions were inspected manually with the Bruker Compass Data Analysis software (Version 4.4) to group the features (precursors and fragments) stemming from the same metabolite in high resolution MS based untargeted experiments. ChemStation software was used in the data analysis in single quadrupole MS based untargeted experiments. The targeted ion extraction window was  $\pm 0.01$  m/z in high resolution MS data analysis and (-0.3, +0.7) da in single quadrupole MS data analysis. The peaks were carefully selected based on the retention time of each metabolite. The MS/MS spectra generated at 20 eV in the METLIN library were used as reference spectra. As stated earlier, two parameters, including ratio score and match factor, were used as indicators to assess the similarity of fragments generated at two different conditions.<sup>7</sup> Ratio score was calculated as a fraction, with the denominator as the total number of fragments in the 20 eV MS/MS spectra in



the METLIN library, unless stated otherwise. The numerator was the number of fragment ions in MS<sup>1</sup> (generated with eMISA or DIA) or MS/MS (generated at DDA mode) that matched to the denominator. Match factor was calculated using the cosine dot product, which has been widely used in the comparison of mass spectral similarity.<sup>22</sup> In this study, we computed the cosine of the angle between two sequences of relative intensities and was used to assess the fragmentation similarity.<sup>22</sup> The match factor was only calculated for those metabolites with 3 or more fragments produced. Match factor ranged from 0 to 100 (%), with 100 being the highest similarity match. Precursors were excluded from the computation of both ratio score and match factor. Further, only those fragments with over 5% relative intensity in the 20 eV MS/MS spectra in the METLIN database were used, unless all the fragments have relative intensities below 5% at 20 eV. The fragments outside the mass acquisition range (below 50 *m/z*) were not considered. Statistical analysis was conducted with GraphPad Prism 8.

## RESULTS AND DISCUSSION

**In-source fragmentation condition optimization.** This study aimed to find in-source fragmentation conditions in which more fragments can be generated without compromising the intensity of precursor ions (Fig. 1). Specifically, the aim is to generate mass spectral data including both the high abundance precursor ions and their corresponding in-source fragments for confident compound identification within a single run in full scan mode. The 20 eV MS/MS spectra in the METLIN library, produced in the collision cell with analytical standards, was used as a reference for the optimization process. We started with optimizing the electrospray ionization (ESI) source parameters such as capillary voltage and dry gas temperature to increase in-source fragmentation. Capillary voltage is known to control the desolvation/activation energy, which is directly related with the in-source fragmentation.<sup>23,24</sup> One purpose of dry gas within the source is to aid desolvation in electrospray source. Higher capillary voltage and dry gas temperature can facilitate the generation of in-source fragments.<sup>23,24</sup> In the BRUKER impact II, the suggested range of capillary voltage is 2000 to 5000 V at positive mode and 1500 to 4000 V at negative mode, respectively; the suggested range of dry gas temperature is 180 to 240 °C, and the capillary voltage was set at 4500 V at positive mode and 4000 V at negative mode, respectively; the dry gas temperature was set up as 180 °C. Thus, we first increased the dry gas temperature to 240 °C to observe the impact on in-source fragmentation. As exemplified with

four metabolites in Table 1, 1-palmitoyl-sn-glycero-3-phosphocholine and cytidine monophosphate in the positive mode and uridine and glucose-6-phosphate in the negative mode, increasing dry gas temperature does not significantly impact in-source fragmentation.

Next, we focused on the optimization of energy between ion funnel 1 and ion funnel 2, called the transfer isCID energy. The MS<sup>1</sup> spectra for the above mentioned four metabolites at two transfer isCID energies (20 and 40 eV) are shown in Table 1, where increasing transfer isCID energy from 0 (original setting) to 40 eV significantly improved the in-source fragmentation of selected molecules at both ionization modes.

To test the generality of this observation and find the appropriate transfer isCID energy for in-source fragmentation, we produced the MS<sup>1</sup> spectra containing in-source fragments for a mixture of 50 endogenous metabolites (30 μM) at varying transfer isCID energies: from 0 to 100 eV were tested in 10 eV increments. These metabolites were all observed at negative mode and their MS/MS spectra at – 20 eV were all available in the METLIN library, however, at positive mode, only 33 metabolites were observed. Three factors were mainly considered in selecting the appropriate ISF condition: the number and relative intensity of fragments that matched to 20 eV MS/MS spectra in the METLIN database and the intensity of precursor ion.

To assess the in-source fragmentation performance, we first calculated the percentage of fragments of each metabolite generated at each condition versus the total number of fragments recorded in the 20 eV MS/MS spectra in the METLIN database. Only those fragments with a relative intensity above 5% were considered. As shown in Fig. 2a and 2b, the median percentage observed at both positive and negative modes increases with the increase of isCID energy. Similar trend was also observed for match factor (Fig. 2c and 2d), which was calculated for those metabolites with at least 3 fragments in the 20 eV MS/MS spectra in the METLIN database to assess the relative intensity similarity. The relationship between the median value of 1<sup>st</sup> top fragment intensity of each metabolite versus the isCID energy exhibited a reversed U-shaped curve (Fig. 2e and 2f), especially in the positive mode. This indicates that there are optimal and consistent settings across a wide variety of molecules for consistently creating fragments that facilitate identification. We further evaluated these setting with respect to intensity loss of precursor ion and as expected, the intensity of the precursor ion decreases with an increase of isCID energy (Fig. 2g and 2h).

Based on the results observed, we selected 40 eV and 30 eV isCID energy as the eMISA condition at negative and positive mode, respectively, with all other parameters fixed. Compared with the original in-source condition (isCID energy is 0 eV), the MS<sup>1</sup> spectra acquired at 40 eV isCID energy in the negative mode shows better fragmentation pattern: 37% increase in median percentage of fragments (up to 100% in over half cases), 15% increase in median match factor, and 3.2 times increase in the median intensity of 1<sup>st</sup> top fragment ion (Fig2a,c,e). Further, as compared with higher isCID energies, a more constant precursor ion intensity (a median loss of only 9%) was observed when isCID energy is 40 eV (Fig. 2g). In the positive mode, MS<sup>1</sup> spectra at 30 eV isCID energy has close median match factor and major fragment ion intensity with the original in-source fragmentation condition, but it has higher median percentage of fragments, 96%, which is close to what was observed at higher isCID energies (100% at 40 eV and higher energies) (Fig. 2b,d,f). More importantly, the median loss of precursor ion intensity at this condition is only 7% lower, however, this number is 47% at 40 eV isCID energy, although significantly improved fragmentation data was achieved (Fig. 2h). These results indicate that it is possible to generate in-source fragments comparable with higher energy MS/MS spectra by enhancing ISF condition in the ESI source, while maintaining the intensity of precursor ion.

**eMISA vs. QTOF DDA mode.** Tandem mass spectra similarity is a crucial factor in compound identification in untargeted experiments. Fragmentation patterns produced by the data dependent acquisition (DDA) mode in high resolution mass spectrometry are considered as the gold standard in compound identification. Here, fragmentation data was acquired for the target molecules using both eMISA and QTOF DDA (collision energy: 20 eV).

The ratio scores were generated to compare the numbers of fragments produced with eMISA mode versus the numbers of fragments generated at DDA mode (20 eV) for each molecule (Table 2). When considering the fragments with relative intensity above 5%, 34 out of 50 molecules at negative mode and 17 out of 33 molecules at positive mode had matching fragments between the two approaches; when considering those fragments with relative intensity above 30%, the numbers increased to 46 out of 50 at negative mode and 30 out of 33 at positive mode, respectively (Fig. 3a). Since major fragment ions ( $\geq 30\%$  relative intensity) are vital in MS/MS fragmentation pattern matching, eMISA demonstrated the capability to generate fragments comparable to the DDA mode for 92% metabolites studied. The similarity of the

fragmentation patterns generated by the two approaches was further evaluated by calculating match factors for those metabolites with over three fragments. As seen in Table 2 and Fig. 3b, the match factors ranged from 10% to 100% (median: 87%) at negative mode and from 14% to 100% (median: 71%) at positive mode, respectively, suggesting good spectra fragment similarity for most molecules between the two modes.

Next, the peak intensity of fragments was compared between the two approaches. Among all the fragments produced at negative mode, 49% of them had higher absolute intensities at eMISA mode with a median of 0.91 times; 67% of fragments generated in the positive mode had higher absolute intensities at eMISA mode with a median of 1.7 times (Fig. 3c). This suggests that the intensity of fragments generated at eMISA mode is similar or higher than that produced in the DDA mode.

Untargeted metabolomics studies often include two sets of experiments, MS<sup>1</sup> metabolite profiling followed by MS/MS spectra generation for compound identification. In MS<sup>1</sup> metabolite profiling, efforts are normally made to minimize the ISF to reduce the number of false positive features from in-source fragmentation, selected features are then analyzed for MS/MS spectra generation. This work indicates that by enhancing the ISF that putative identifications can be made with higher confidence based on in-source fragmentation patterns consistent with that generated in MS/MS experiments at 20 eV collision energy, without compromising the intensity of precursor ions.

It is worth noting that in DDA the absolute intensities of most precursor ions dropped significantly at higher collision energies in contrast to eMISA. Overall, all the precursor ions observed at eMISA mode had much higher absolute intensities than their counterparts acquired in the DDA mode with a median of 18 times at negative mode and 210 times at positive mode, respectively (Fig. S1). The quantitative relationship of the intensity of both precursor ions and fragments generated between eMISA and DDA modes was exemplified with four metabolites, phenylalanine and tryptophan in the positive mode and fructose-6-phosphate and oxidized glutathione in the negative mode (Fig. 4). This high abundance precursor ion intensity for eMISA helps us understand why fragmentation generation is possible without significantly compromising sensitivity as compared to DDA. The high abundance of precursor ions in the eMISA MS<sup>1</sup> spectra help trace the origins of the generated fragments.

A useful feature of eMISA mode is that fragment ions can be generated across a broad dynamic range of ion intensities. However, a low intensity of precursor ions can compromise DDA MS/MS fragmentation pattern generation. For the investigation of sensitivity, a mixture of four metabolites at each ionization mode was analyzed by eMISA and DDA at 9 concentrations ranging 6 orders of magnitude: 1 nM, 10 nM, 100 nM, 1  $\mu$ M, 10  $\mu$ M, 50  $\mu$ M, 100  $\mu$ M, 500  $\mu$ M, and 1 mM (see Methods for details). For example, in the DDA analysis at negative mode, two fragments of tyrosine with a relative intensity of 42% and 27%, respectively, are missing at 10  $\mu$ M and many interference fragments with high abundance were still present at this level, which could be problematic in the identification process (Table S2 and Fig. S2). However, eMISA generated two major fragments of tyrosine, with relative intensity of 100% and 90%, respectively, even at 1 nM with a high abundance of precursor ion formation (Table S2). Further, we calculated the limit of detection (LOD) for each metabolite with the LOD being determined as the precursor ion intensity being 3 times the signal/noise ratio. As shown in Table 3, eMISA exhibited one to two magnitudes lower LOD beyond DDA.

Overall, compared with QTOF DDA based MS/MS approaches, eMISA can generate the full scan mass spectra ( $MS^1$ ) containing both the precursor ions and their fragments within a single run. It enables the maximum information collection even for analytes at low levels which may be missed if using the targeted MS/MS technique. The high intensities of precursor ions observed in the  $MS^1$  generated with eMISA increases sensitivity and confidence in compound identification. By delivering one to two orders of magnitude deeper sensitivity, eMISA technique may enable the application of high-resolution mass spectrometry in full scan mode for direct low abundant metabolite identification and relative quantification.

**eMISA and QTOF DIA.** Similar to eMISA, data independent acquisition (DIA) with high-resolution mass spectrometry does not require the initial detection of MS peaks to proceed to MS/MS analysis. In the BRUKER impact II system, the DIA technique is called Broadband Collision Induced Dissociation (bbCID). No precursor ion isolation is required in DIA and all ions are fragmented in the collision cell resulting in the generation of DIA MS/MS spectra across every LC peak. The workflows for the two techniques are illustrated in Fig. 5a and 5b, respectively.

We first compared the mass spectra of the select molecules generated by the two approaches. The number of fragments generated with eMISA that matched to 20 eV MS/MS spectra generated in DIA were calculated as a ratio score. As shown in Table S1, 33 in 50 metabolites had all the fragments matched between the two techniques in the negative mode; in the positive mode, 16 of 33 metabolites had a complete match between the two approaches. In general, over 90% molecules at both polarities had at least half the number of the fragments produced at eMISA, compared to the DIA (Table S1). We then calculated the match factors to assess the similarity of the fragment spectra generated by the two techniques. In the negative mode, the median match factor is 82% and 68% molecules had a match factor over 60%; separately in the positive mode, the median match factor is 81% and 86% molecules had a match factor over 60% (Table S1; Fig. 5c). This suggests a good match for the fragmentation patterns produced by the two techniques.

To explore the advantages of eMISA over DIA, we compared the absolute intensities of the precursor ions and fragments for all the metabolites produced using the two techniques. In the negative mode, an increase of the absolute intensity was observed in over 60% precursor ions acquired with eMISA at a median of 1.2 times; in the positive mode, the increase was observed in 88% precursor ions at a median of 1.8 times (Fig. 5d). This indicates that precursor ions generated with eMISA has equal or higher intensity opposed to that acquired with DIA. We further investigated the precursor ion sensitivity of the two approaches with the mixture of four metabolites at each polarity mentioned earlier and eMISA demonstrated equal or better sensitivity with the tested metabolites (Table 3). This improvement observed with eMISA can be explained by its unique fragmentation mechanism. DIA fragmentation in the collision cell stops when the duty cycle time ends. Because high abundance analytes occupy a significant amount of cycle time, however low abundant analytes may be not fragmented within a reasonable time frame, thus limiting their availability for data collection. However, eMISA fragments every peak simultaneously without the limit of cycle time. For the fragments, most peaks produced in both negative (92%) and positive (86%) ionization modes are higher with DIA by a median of 4.3 and 4.6 times, respectively (Fig. 5e).

Overall, instead of submitting the MS<sup>1</sup> with many in-source fragments to the collision cell for MS/MS acquisition as in DIA, eMISA produces the pseudo MS/MS spectra for molecules

comparable to the 20 eV MS/MS spectra produced in the DIA mode. This technique demonstrates equal or better sensitivity of eMISA for the studied molecules.

**Application of eMISA for metabolite annotation in untargeted metabolomics.** To illustrate the capacity of eMISA in compound identification in untargeted metabolomics, we analyzed metabolites from a murine macrophage-like cell line (RAW264.7) in the enhanced ISF condition using high resolution mass spectrometry (QTOF-MS). eMISA acquired the MS<sup>1</sup> data containing all the molecular ions in the sample and their fragment ions. Table 4 shows the list of correctly annotated metabolites using eMISA. Thirteen compounds were identified in both positive and negative modes (Table 4). Of the 50 molecules studied, we successfully identified the presence of 41 metabolites in the macrophage extracts based on two indicators, ratio score and match factor, by matching to the 20 eV MS/MS spectra in the METLIN database (Table 4). eMISA also identified the presence of metabolites in the macrophages extract beyond the list, such as itaconate (Table 4). The identifications were confirmed by the analysis of analytical standards.

**eMISA and compound identification with an ESI single quadrupole MS.** The striking similarity between ESI in-source fragments of small molecules and MS/MS data may enable untargeted experiments on more general (simple) MS platforms, such as single quadrupole instruments. To examine this possibility, we analyzed the mixture of 50 molecules using an LC ESI single quadrupole MS at two fragmentor voltage, one of the key parameter settings in tuning the instrument's in-source fragmentation energy. Compounds with easily fragmented chemical structures (e.g., amino acids) typically fragment at low fragmentor voltages while others (e.g., lipids) fragment at higher voltages, higher voltages may preclude the observation of precursor ions for the more fragile molecules. To compensate for variability in structural vulnerability to fragmentor voltage we selected two fragmentor voltages, 150 and 300 V, to produce in-source fragments for those compounds requiring low and high collision energies, respectively. Overall, in-source fragment information consistent with METLIN data was successfully acquired for 36 of the 50 molecules (30 in positive mode and 36 in negative mode). To identify the molecule, the number of hits against METLIN library ranged from 1 to 305 with a median of 34 (n=66) when only using the precursor ion information (Fig. 6). However, when the in-source fragments information was also considered, the number of hits significantly dropped to a level comparable

with that obtained using high resolution mass spectrometry, as shown in Fig. 6. Our results demonstrated that eMISA strategy enabled single quadrupole MS based untargeted LC/MS datasets to be more useful.

It should be noted that the ESI single quadrupole used in these experiments is a much lower sensitivity instrument with a limited dynamic range as compared to the QTOF used in the previous studies. However, and similar to many database-based untargeted LC-MS workflows, compounds can be identified with eMISA are limited to the compounds included in the METLIN MS/MS spectral database. For the identification of truly unknown compounds, targeted MS/MS analysis is still required to produce reliable MS/MS spectra for structural confirmation. Further, only Bruker QTOF instrument and Agilent single quadrupole MS were examined for the validation of eMISA strategy in this study.

## CONCLUSIONS

LC-MS experiments are typically designed to minimize in-source fragmentation, however, this study demonstrates that enhancing ISF can be used to improve both sensitivity and confidence in compound identification from MS<sup>1</sup> experiments. Enhanced ISF can provide in-source fragments similar to 20 eV MS/MS spectra in the METLIN library in terms of intensity and number of fragments, thus allowing a straightforward and robust metabolite annotation to significantly facilitate subsequent metabolite identification. METLIN currently has the experimental MS/MS spectra for over 700,000 molecules, making eMISA a promising approach for the autonomous and robust annotation and identification of a broad range of metabolites and small molecules.

Compared with the other DDA and DIA approaches, higher sensitivity ranging from one to two orders of magnitude were observed over DDA while eMISA provided equal or higher sensitivity as compared to DIA. By using a metabolite-centric approach, eMISA doesn't lose the direct link between a precursor ion and their fragments, which is commonly lost in DIA data analysis. Further to this, since collision cell is not a prerequisite in fragments generation, eMISA enables MS<sup>1</sup> data acquired on ESI time-of-flight and quadrupole instruments (single and triple) to be more useful for molecular identification. This last point is particularly compelling, as eMISA will allow for untargeted LC/MS experiments to be performed with a significantly higher



level of annotation confidence to be performed on the tens of thousands of quadrupole instruments currently relegated to MS<sup>1</sup> experiments.

## REFERENCES

- (1) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. *Nat Rev Mol Cell Biol* **2016**, *17*, 451-459.
- (2) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal Chem* **2011**, *83*, 2152-2161.
- (3) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. *Anal Chem* **2018**, *90*, 480-489.
- (4) 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. *Anal Chem* **2018**, *90*, 3156-3164.
- (5) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. *Mol Biosyst* **2012**, *8*, 470-481.
- (6) Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimera, R.; Sales-Pardo, M. *Bioinformatics* **2019**, *35*, 4089-4097.
- (7) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Guijas, C.; Majumder, E. L.; Benton, H. P.; Siuzdak, G. *Anal Chem* **2019**, *91*, 3246-3253.
- (8) Mahieu, N. G.; Patti, G. J. *Anal Chem* **2017**, *89*, 10397-10406.
- (9) Lu, W.; Su, X.; Klein, M. S.; Lewis, I. A.; Fiehn, O.; Rabinowitz, J. D. *Annu Rev Biochem* **2017**, *86*, 277-304.
- (10) Abranko, L.; Garcia-Reyes, J. F.; Molina-Diaz, A. *J Mass Spectrom* **2011**, *46*, 478-488.
- (11) Pais, P.; Moyano, E.; Puignou, L.; Galceran, M. T. *J Chromatogr A* **1997**, *775*, 125-136.
- (12) Tian, Q.; Duncan, C. J.; Schwartz, S. J. *J Mass Spectrom* **2003**, *38*, 990-995.
- (13) Wickens, J. R.; Sleeman, R.; Keely, B. J. *Rapid Commun Mass Spectrom* **2006**, *20*, 473-480.
- (14) Xu, Y. F.; Lu, W.; Rabinowitz, J. D. *Anal Chem* **2015**, *87*, 2273-2281.
- (15) Gathungu, R. M.; Larrea, P.; Sniatynski, M. J.; Marur, V. R.; Bowden, J. A.; Koelmel, J. P.; Starke-Reed, P.; Hubbard, V. S.; Kristal, B. S. *Anal Chem* **2018**, *90*, 13523-13532.
- (16) Purwaha, P.; Silva, L. P.; Hawke, D. H.; Weinstein, J. N.; Lorenzi, P. L. *Anal Chem* **2014**, *86*, 5633-5637.
- (17) Wu, J. L.; Liu, J.; Cai, Z. *Rapid Commun Mass Spectrom* **2010**, *24*, 1828-1834.
- (18) Carrier, D. J.; Eckers, C.; Wolff, J. C. *J Pharm Biomed Anal* **2008**, *47*, 731-737.
- (19) Seitzer, P. M.; Searle, B. C. *J Proteome Res* **2019**, *18*, 791-796.
- (20) Li, H.; Cai, Y.; Guo, Y.; Chen, F.; Zhu, Z. J. *Anal Chem* **2016**, *88*, 8757-8764.
- (21) Warth, B.; Raffener, P.; Granados, A.; Huan, T.; Fang, M.; Forsberg, E. M.; Benton, H. P.; Goetz, L.; Johnson, C. H.; Siuzdak, G. *Cell Chem Biol* **2018**, *25*, 291-300 e293.
- (22) Kim, S.; Zhang, X. *Comput Math Methods Med* **2013**, *2013*, 509761.
- (23) Purvine, S.; Eppel, J. T.; Yi, E. C.; Goodlett, D. R. *Proteomics* **2003**, *3*, 847-850.
- (24) Ramos, A. A.; Yang, H.; Rosen, L. E.; Yao, X. *Anal Chem* **2006**, *78*, 6391-6397.

Table 1. Absolute (relative) intensities comparison (including both precursor ions and fragments) for metabolites across different in-source fragmentation conditions and METLIN library at both positive mode (1-palmitoyl-sn-glycero-3-phosphocholine and cytidine monophosphate) and negative mode (glucose-6-phosphate and uridine)

analyte	m/z	absolute (relative) intensity					relative intensity
		dry gas temperature (°C)		isCID energy (eV)			METLIN (20 eV)
		180	240	0	20	40	
1-palmitoyl-sn-glycero-3-phosphocholine (ESI(+))							
precursor ion	496.34	9896886	9789667	9896886	9289506	8081574	
fragment 1	104.105	0	0	0	0	20488 (100)	54
fragment 2	184.069	0	0	0	0	11562 (56)	100
fragment 3	478.319	14480	15260	14480	11520	13658 (67)	10
cytidine monophosphate (ESI(+))							
precursor ion	324.053	133396	145672	133396	139096	86498	
fragment 1	97.029	5142	5230	5142	7032	15258	5
fragment 2	112.048	1814604	1867800	1814604	2120736	4135146	100
glucose-6-phosphate (ESI(-))							
precursor ion	259.022	454652	468930	454652	475116	329692	
fragment 1	78.959	1672 (2)	1749 (2)	1672 (2)	2610 (3)	14110 (30)	38
fragment 2	96.9693	57424 (60)	56390 (59)	57424 (60)	49550 (52)	47024 (100)	100
fragment 3	138.978	95464 (100)	96480 (100)	95464 (100)	95368 (100)	24776 (53)	8
uridine (ESI(-))							
precursor ion	243.062	3674204	3756890	3674204	3928470	3615684	
fragment 1	66.0347	0		0	0	7134 (3)	15
fragment 2	82.0301	0		0	0	18696 (8)	37
fragment 3	110.025	41574 (100)	42750 (100)	41574 (100)	40960 (100)	242978 (100)	100
fragment 4	122.024	0		0	0	14958 (6)	11
fragment 5	152.034	16796 (40)	17840 (42)	16796 (40)	21004 (51)	75444 (31)	8
fragment 6	153.03	13222 (32)	14630 (34)	13222 (32)	18756 (46)	62994 (26)	6

Table 2. Fragmentation pattern comparison between mass spectra acquired with eMISA mode and QTOF DDA mode (20 eV)

no.	ESI(-)			ESI(+)		
	name	score <sup>a</sup>	M.F. <sup>b</sup>	name	score	M.F.
1	aspartate	4/4	95			
2	malate	4/4	80			
3	adenine	2/4	100	adenine	2/3	100
4	hypoxanthine	2/3	100	hypoxanthine	2/5	95
5	sulfoacetic acid	2/2				
6	glutamine	8/9	84	glutamine	3/3	28
7	glutamate	2/2		glutamate	3/3	63
8	methionine	1/1		methionine	6/7	42
9	xanthine	2/2*	100	xanthine	1/4	99
10	aminoadipic acid	5/6*	96	aminoadipic acid	4/5	57
11	phenylalanine	4/5	87	phenylalanine	2/2	
12	phosphoenol pyruvate	1/1				
13	uric acid	4/4	100			
14	glyceraldehyde-3P	2/2				
15	arginine	1/1		arginine	4/4	41
16	tyrosine	4/4	88	tyrosine	8/9	73
17	acetyl-glutamic acid	5/6	87	acetyl-glutamic acid	4/4	89
18	citric acid	5/5	90			
19	tryptophan	4/4	94	tryptophan	7/7	48
20	myristic acid	0/0				
21	ribose-5-phosphate	3/3	63			
22	cytidine	5/5	100	cytidine	1/1	
23	uridine	6/6	98	uridine	1/3	100
24	palmitic acid	0/0		palmitic acid	12/12	62
25	glucose-6-phosphate	3/3	94	glucose-6-phosphate	5/5*	97
26	fructose-6-phosphate	3/3	94	fructose-6-phosphate	2/2	
27	adenosine	1/1		adenosine	1/1	
28	inosine	1/1		inosine	1/1	
29	saccharopine	7/14	97	saccharopine	2/4	68
30	linoleic acid	0/0				
31	oleic acid	0/0		oleic acid	23/25	77
32	guanosine	3/3	100	guanosine	1/1	
33	EPA	1/7	94			
34	CMP	3/3	71	CMP	2/2	
35	UMP	6/6	73	UMP	2/2	
36	fructose-1,6-bisphosphate	5/5	91			
37	AMP	3/3	71	AMP	1/1	
38	IMP	3/5	81	IMP	1/2	
39	GMP	3/3	55	GMP	1/2	
40	UDP	4/9	17			
41	ADP	5/7	25	ADP	2/2*	
42	GDP	4/6	20	GDP	1/1	
43	PG(16:0/0:0)	1/3	99	PG(16:0/0:0)	2/3	100
44	lysoPC(16:0)	1/2*	100	lysoPC(16:0)	1/3	14
45	GTP	8/8*	23			
46	UDP-galactose	3/4*	10			
47	UDP-N-acetyl-glucosamine	5/8	50			
48	oxidized glutathione	9/9*	60	oxidized glutathione	5/9	96
49	dioleoylphosphatidic acid	3/3*	55			
50	acetyl-CoA	4/4	64	acetyl-CoA	4/6	49

<sup>a</sup>score: ratio score; <sup>b</sup>M.F.: match factor; \*: the number of fragments found in DDA mode is less than that recorded in the 20 eV mass spectra in the METLIN database. Calculation of both score and M.F. was based on the mass spectra produced in QTOF DDA mode.

Table 3. Instrumental sensitivity comparison of the quadrupole time of flight mass spectrometry at three different modes with select metabolites.

analyte	mode	target ion	RT <sup>a</sup> (min)	LOD <sup>b</sup> (nM)		
				eMISA	QTOF DIA	QTOF DDA
tyrosine	ESI(-)	180.0666	17.3	1	1	100
inosine	ESI(-)	267.0735	16.6	1	1	1
UMP	ESI(-)	323.0286	31.8	1	1	1
Oxidized glutathione	ESI(-)	611.1447	32.4	1	1	1
glutamine	ESI(+)	147.0764	6.9	1000	1000	10000
histidine	ESI(+)	156.0766	6.7	1	1	100
phenylalanine	ESI(+)	166.0863	7.7	1	1	100
oxidized glutathione	ESI(+)	613.1592	7.1	10	100	100

<sup>a</sup>RT: retention time; <sup>b</sup>LOD: limit of detection.

Table 4. List of correctly identified molecules in a macrophage extract with the MS<sup>1</sup> spectra acquired with eMISA technique

no.	ESI(-)			ESI(+)		
	name	score <sup>a</sup>	M.F. <sup>b</sup>	name	score	M.F.
1	itaconic acid	1/1				
2	aspartate	3/4	82			
3	malate	4/4	35			
4	adenine	2/4	97			
5	hypoxanthine	1/3	92			
6	sulfoacetic acid	1/2				
7	glutamine	7/9	55	glutamine	3/3	25
8	glutamate	2/2		glutamate	3/3	71
9	methionine	1/1		methionine	5/7	10
10	xanthine	1/3	99			
11	aminoadipic acid	5/7	74			
12	phenylalanine	4/5	39	phenylalanine	2/2	
13	phosphoenol pyruvate	1/1				
14	uric acid	4/4	78			
15	arginine	1/1		arginine	2/4	20
16	tyrosine	3/4	91	tyrosine	8/9	61
17	citric acid	5/5	73			
18	tryptophan	4/4	48	tryptophan	7/7	22
19	ribose-5-phosphate	3/3	80			
20	cytidine	2/5	65			
21	uridine	6/6	93			
22	palmitic acid	0/0		palmitic acid	12/12	57
23	inosine	1/1		inosine	1/1	
24	saccharopine	7/14	10	saccharopine	3/4	55
25	linoleic acid	0/0				
26	oleic acid	0/0		oleic acid	24/25	27
27	EPA	3/7	93			
28	CMP	3/3	66			
29	UMP	6/6	64			
30	fructose-1,6-bisphosphate	5/5	26			
31	AMP	3/3	52			
32	IMP	4/5	78			
33	GMP	3/3	61			
34	UDP	6/9	72			
35	ADP	4/7	64			
36	GDP	4/6	42			
37	PG(16:0/0:0)	1/3	98	PG(16:0/0:0)	3/3	100
38	GTP	8/10	32			
39	UDP-galactose	4/6	25			
40	UDP-N-acetyl-glucosamine	4/8	52			
41	oxidized glutathione	8/10	69	oxidized glutathione	5/9	85
42	acetyl-CoA	2/4	67			

<sup>a</sup>score: ratio score; <sup>b</sup>M.F.: match factor. Calculation of both ratio score and match factor were based on the 20 eV MS/MS spectra in the METLIN library.

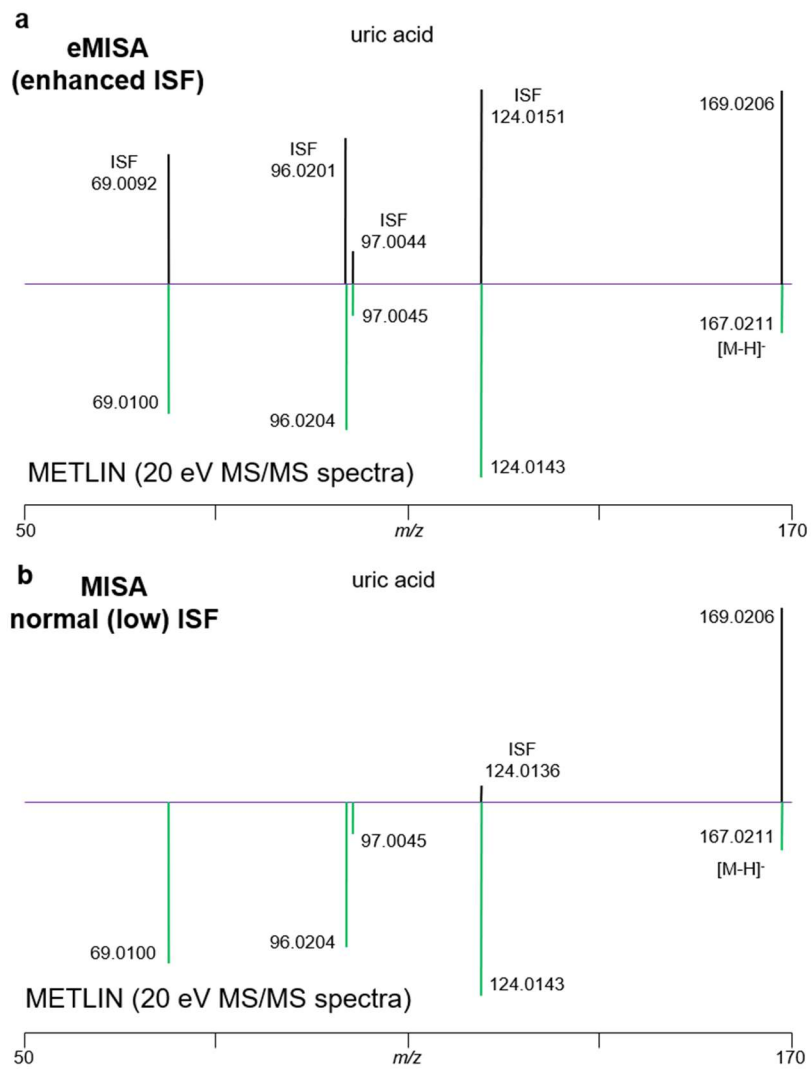


Figure 1. eMISA enhanced ISF fragmentation pattern compared to both non-enhanced ISF as well as METLIN's 20 eV MS/MS spectra for uric acid in the negative mode (ESI(-)).

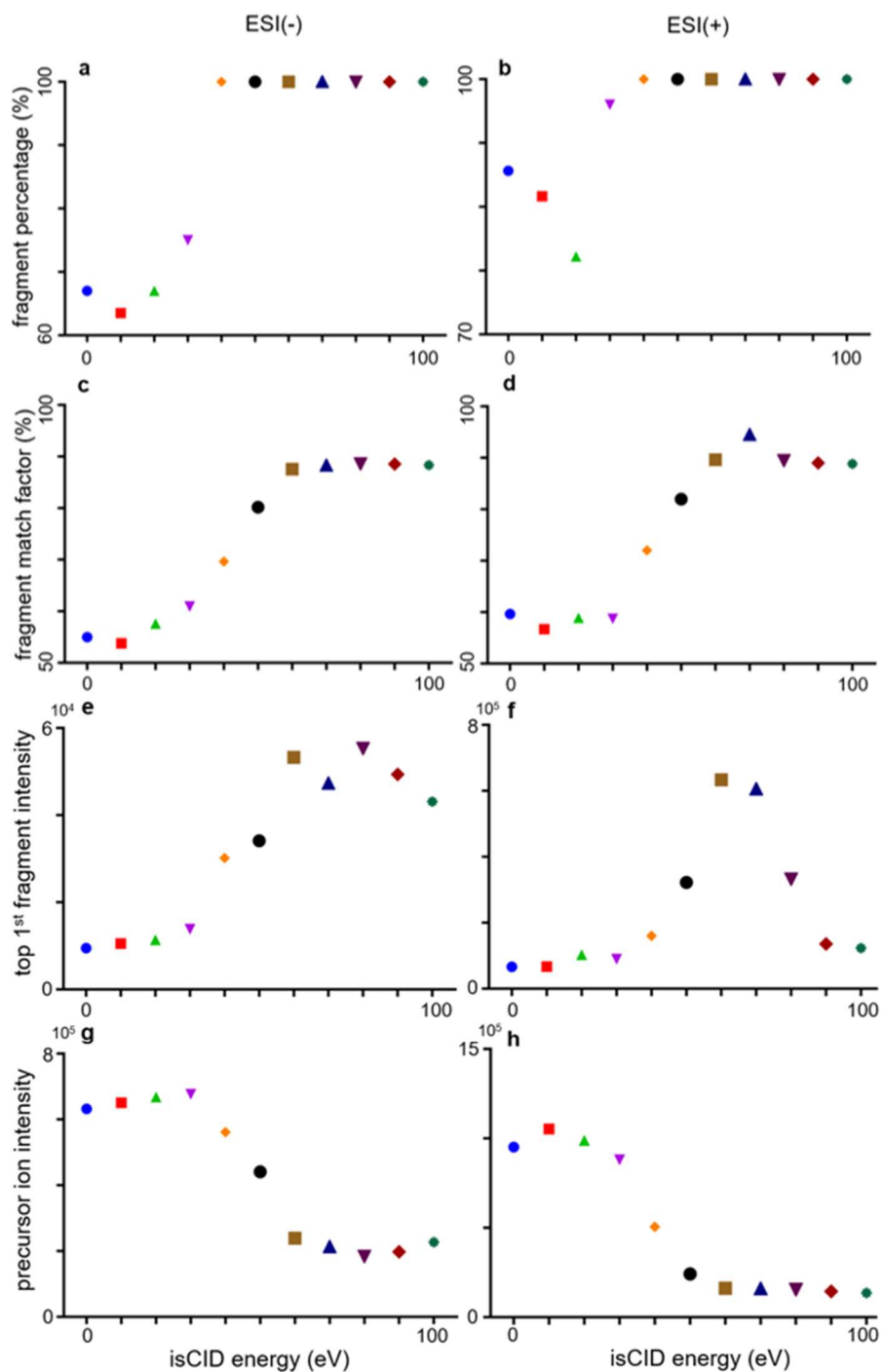


Figure 2. Panels (a) and (b) plot the relationship between median fragment percentage and isCID energy. Panels (c) and (d) plot the relationship between median fragment match factor versus isCID energy. Panels (e) and (f) plot the relationship between median intensity of the major (1<sup>st</sup> top) fragment ion versus isCID energy. Panels (g) and (h) plot the relationship between median intensity of the precursor ion versus isCID energy. Panels (a), (c), (e), and (g) were observed at negative mode (ESI(-)) and others were observed at positive mode (ESI(+)). Both fragment percentage and match factor were calculated versus the +/- 20 eV MS/MS spectra in the METLIN database.

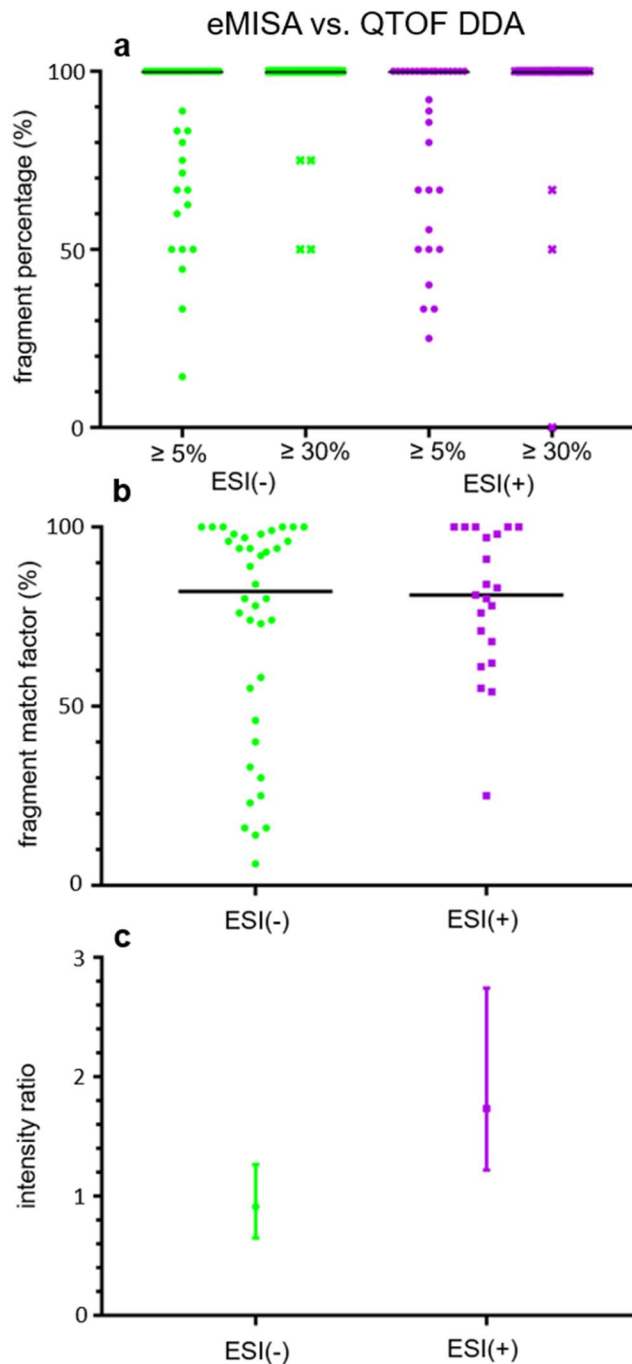


Figure 3. The percentage of fragment produced at eMISA mode versus QTOF DDA mode increases with the increase of relative intensity of fragments (from  $\geq 5\%$  to  $\geq 30\%$ ) considered at both ionization modes with black line representing media (a). Panel (b) shows the vertical scatter plot with median fragment match factor (eMISA vs. QTOF DDA) across different polarity modes. Panel (c) shows the absolute intensity ratio (median plus 95% CI) of fragments generated by eMISA versus those generated with QTOF DDA technique. MS/MS spectra in the QTOF DDA mode was generated at 20 eV.



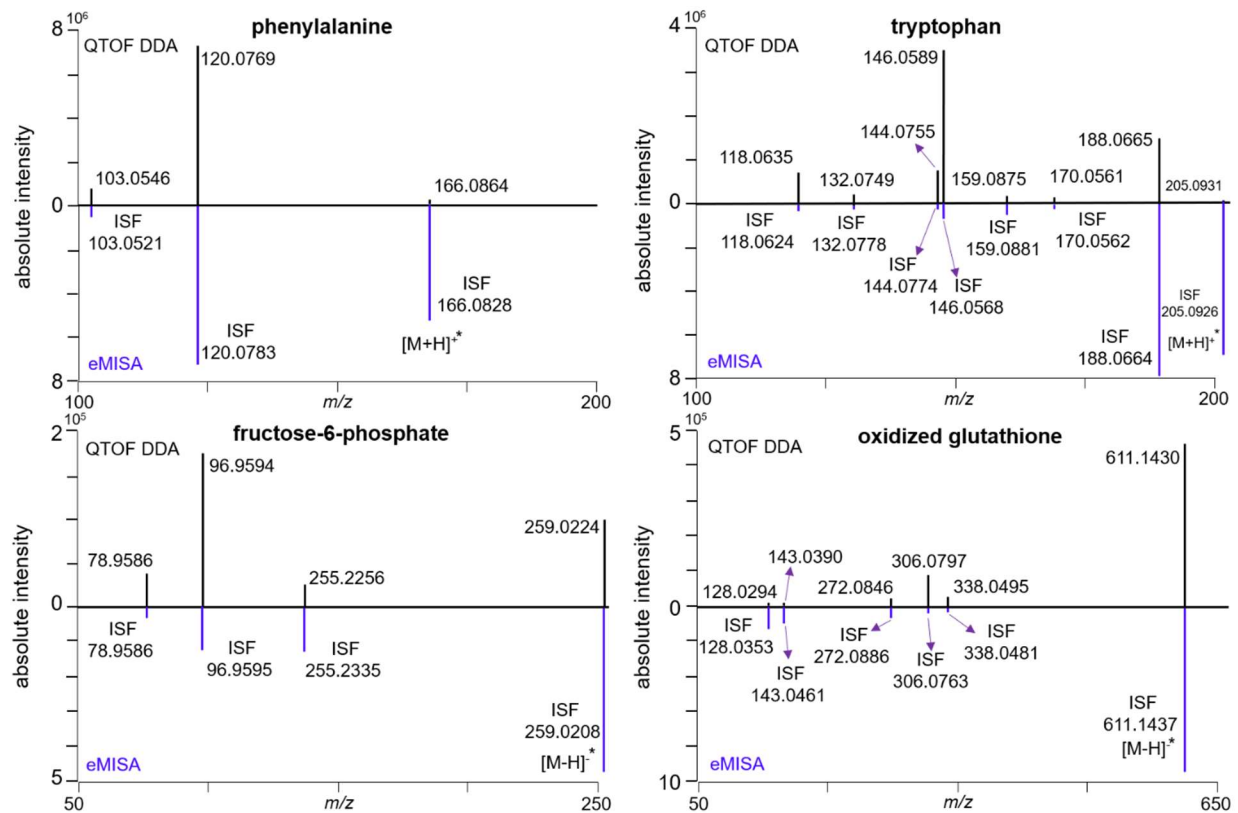


Figure 4. Absolute peak intensity comparison for the precursor ions and the corresponding fragments of four metabolites (phenylalanine and tryptophan at positive mode; fructose-6-phosphate and oxidized glutathione at negative mode) acquired between QTOF DDA mode (20 eV) and eMISA mode.

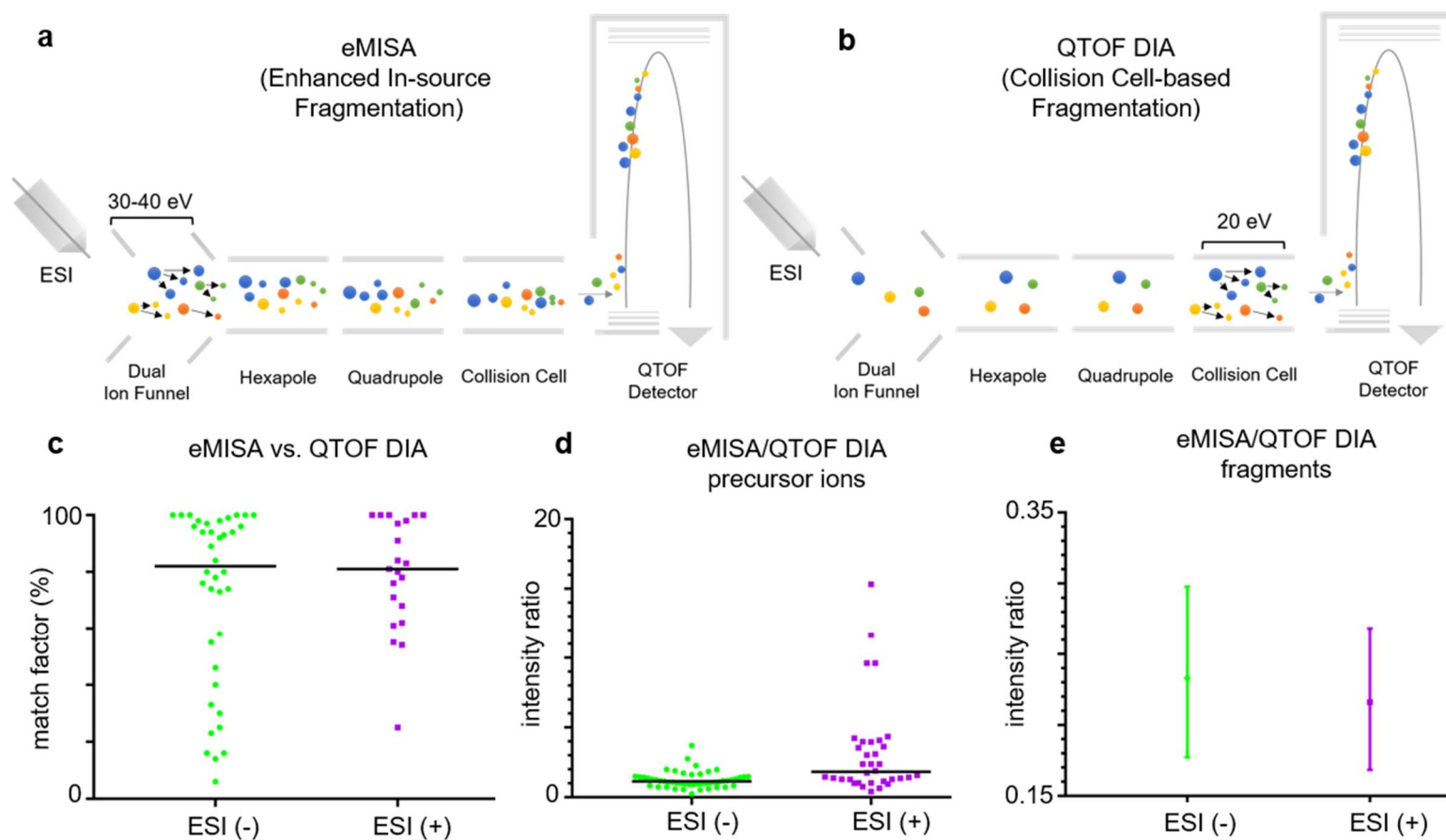


Figure 5. Molecular in-source and collision cell dissociation are provided in panel (a) and (b) for eMISA and QTOF DIA, respectively. Panel (c) shows the vertical scatter plot with median match factor (eMISA vs. QTOF DIA) across different ionization modes. Panel (d) shows eMISA precursor ion intensity/QTOF DIA precursor ion intensity ratios in different ionization modes with black line representing median. Panel (e) shows the absolute intensity ratio of fragments (median plus 95% CI) between eMISA and QTOF DIA technique. MS/MS spectra in the QTOF DIA mode was generated at 20 eV.

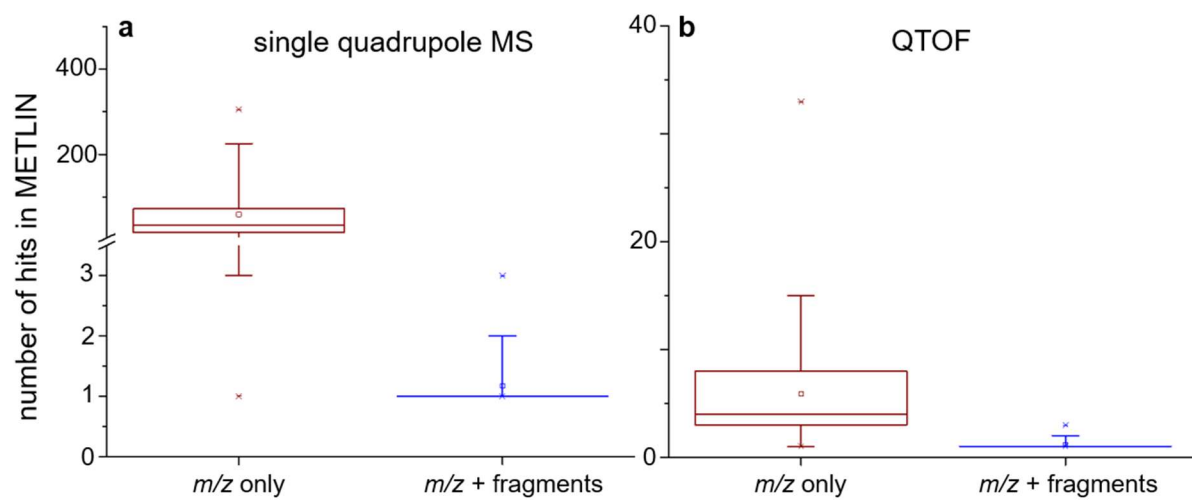


Figure 6. The number of hits obtained in the METLIN library when searching (a) precursor  $m/z$  as well as precursor plus fragmentation (eMISA) data acquired from single quadrupole MS and (b) precursor  $m/z$  as well as precursor plus fragmentation (eMISA) data acquired from QTOF (n=66).