Rapid Communications in Mass Spectrometry **WILEY**

Resolving isobaric interferences in direct infusion tandem mass spectrometry

Jérôme Kaeslin | Renato Zenobi [ID]

Department of Chemistry and Applied Biosciences, ETH Zürich, Zürich, Switzerland

Correspondence
R. Zenobi, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 3, CH-8093 Zürich, Switzerland.
Email: zenobi@org.chem.ethz.ch

Funding information
scholarship fund of the Swiss Chemical Industry (SSCI); ETHZ Foundation

Rationale: The co-fragmentation of precursors in direct infusion (DI) tandem high-resolution mass spectrometry (HRMS) can complicate the fragment spectra and consequently lead to false hits during compound identification.

Methods: The method herein described, termed IQAROS (incremental quadrupole acquisition to resolve overlapping spectra), modulates the intensities of precursors and fragments by stepwise movement of the quadrupole isolation window over the mass-to-charge (m/z) range of the precursors. The modulated signals are then deconvoluted by a linear regression model to reconstruct the fragment spectra with less interference. The hardware to demonstrate the use of IQAROS was an orbitrap with electrospray ionization (ESI) or secondary electrospray ionization (SESI), although the method can also be applied to other ionization techniques or mass analyzers.

Results: Assessing the performance of IQAROS with isobaric standards revealed that the reconstructed fragment spectra match with spectra acquired from the pure standards and that more compounds were correctly identified compared with the classical approach with the quadrupole centered at the m/z value of the precursor of interest. Moreover, the strength of IQAROS is exemplified by the identification of two isobaric biomarkers directly from a breath sample with SESI-HRMS.

Conclusions: With IQAROS, cleaner fragment spectra of co-fragmenting isobars during DI-HRMS analysis can be obtained. IQAROS can easily be set up by the standard graphical user interface of the instrument. Therefore, it facilitates the characterization of features of interest in samples analyzed by DI-HRMS, for example, in high-throughput or real-time metabolomics.

1 | INTRODUCTION

Given its speed and sensitivity, direct infusion (DI) mass spectrometry (MS) is the ideal technique for high-throughput analysis[1] and for real-time on-line monitoring.[2] For example, DI-MS-based metabolomics[3] or on-line MS-based breath analysis[4] are applications where speed and high time resolution are crucial. The excellent separation of ions according to their mass-to-charge ratio (m/z) with high-resolution MS (HRMS) has proven to compensate at least partially for the missing

chromatographic separation.[5,6] In a typical DI-HRMS metabolomics workflow, samples are analyzed on an MS[1] level, followed by statistical analysis which yields MS[1] signals of interest, which in turn have to be annotated to chemical structures in the next step.[4] For sufficient confidence, this identification step requires characterization by tandem mass spectrometry (MS[2]).[7]

While molecules with differences of a few millidaltons can be distinguished on the MS[1] level with HRMS, their MS[2] characterization is more challenging because the quadrupole (Q) isolation window for

precursor selection typically has a width of a few hundred millidaltons to a few daltons. Thus, isobaric precursors are often co-fragmented and interfere with each another's MS[2] spectra. In the literature, the resulting complicated spectrum is often referred to as a chimeric MS[2] spectrum or chimera.[8] If a feature of interest makes up ≤50% of the total MS[1] intensity in the isolation window, a rule of thumb says that spectral matching is unrealiable.[9] Thus, for DI-MS metabolomics, where hundreds of features are detected in the low-mass region,[10,11] spectral matching is often problematic.

This problem is well known in liquid chromatography/MS (LC/MS)-based proteomics or metabolomics, where complex samples often lead to chimeras – even with the chromatographic separation and both for data-independent acquisition (DIA) as well as for precursor-dependent acquisition. Fortunately, several strategies to tackle this problem were developed: (1) re-analyzing the sample under different LC conditions; (2) using the retention time correlation of precursors and fragments to deconvolute and reconstruct MS[2] spectra;[12–17] (3) deconvolute a chimera directly into a linear combination of library MS[2] spectra;[18–20] or (4) use of *in-silico*

fragmentation,[21–25] which ultimately disentangles chimeras by focusing on fragmentations with reasonable mass losses.

For DI-MS, options (1) and (2) are not feasible because they rely on chromatography. Option (3) could be applied to DI-MS, but it assumes that the interfering isobars are reported in a spectral library. However, this might often not be the case for adducts and in-source fragments[26–30] or for ionization techniques other than ESI. Option (4) can have similar limitations, but is still very helpful because the focus on reasonable mass losses is independent of the ionization technique. Nonetheless, it might be that a fragment is consistent with two isobars simultaneously from a pure mass loss perspective, e.g., the benzyl fragment $C_6H_5^+$ could arise from either one (or both) of two co-fragmented aromatic precursors. Only a few studies have explicitly addressed the problem for DI-MS, e.g., it is reported that ion mobility[31] or "*collisional purification*"[32,33] on MS[3]- or pseudo-MS[3]-capable instruments can resolve chimeras.

Here, we report an alternative method dubbed IQAROS (incremental quadrupole acquisition to resolve overlapping spectra) to resolve co-fragmented precursors in DI-HRMS. It relies on small,
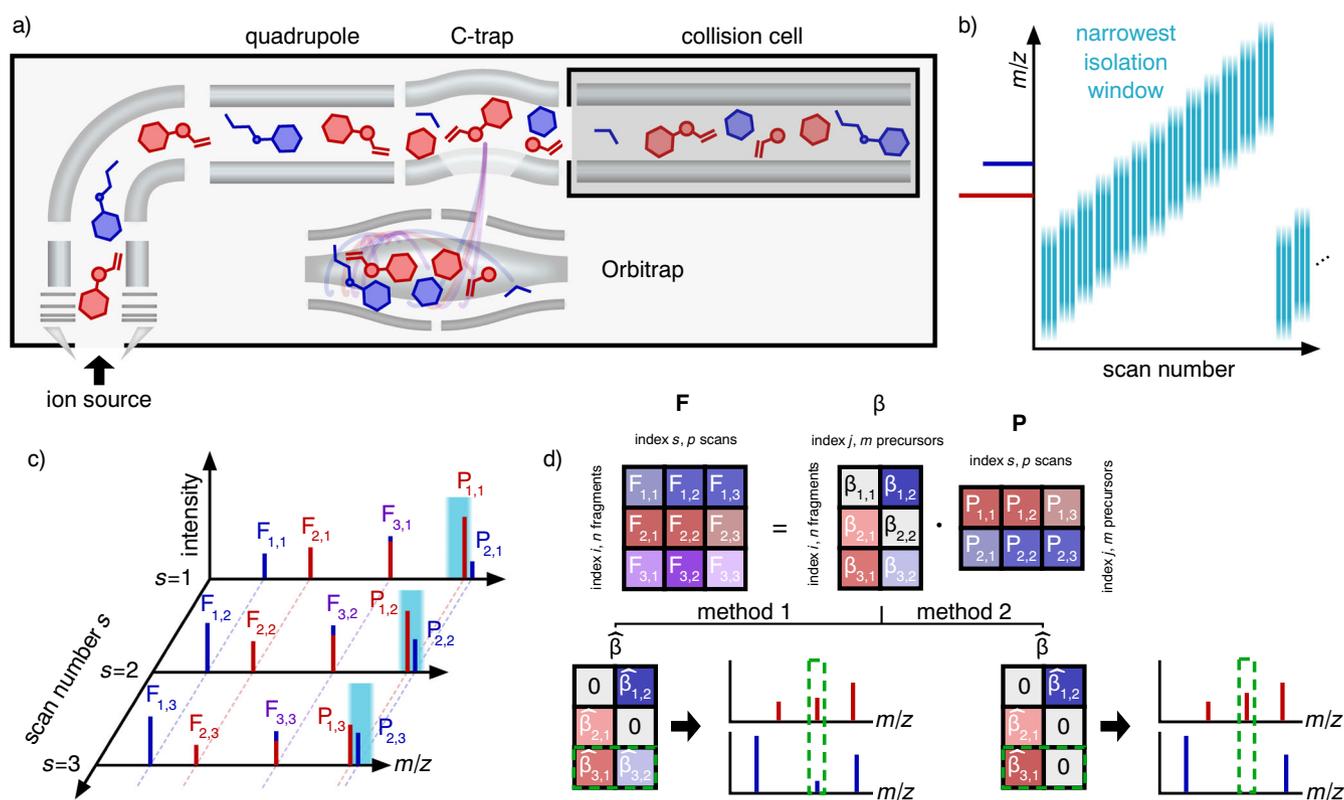
FIGURE 1 Overview of IQAROS: a, Two co-isolated, co-fragmented and co-detected isobars result in a chimeric MS[2] spectrum. b, The basic principle of IQAROS. The isolation window is stepwise moved over the *m/z* range of the two isobaric precursors. On the narrowest isolation width setting, the isobars cannot be isolated individually, but they can be modulated distinctively. c, Moving the isolation window through the precursor region modulates the intensities of the two isobaric precursors. Their scan *s* dependent intensities are denoted $P_{1,s}$ and $P_{2,s}$. Likewise, the fragment intensities are also modulated. d, The observed fragment intensity $F_{i,s}$ of fragment *i* in scan *s* can be expressed by a linear combination of the precursor intensities $P_{j,s}$ in the same scan *s* and a contribution coefficient $\beta_{i,j}$. The estimates $\hat{\beta}$ are obtained with two methods here: a non-negative multi-linear regression (method 1) or a non-negative simple-linear regression (method 2). The green rectangle highlights the difference between the two methods [Color figure can be viewed at wileyonlinelibrary.com]

millidalton differences between the accurate masses of the precursors. The method is based upon the fact that the width of the Q isolation window is relatively broad, but the window's center can be set quite precisely. Thus, by moving the Q over the precursor range in a stepwise manner, the precursor's transmission through the Q is regulated and the intensities of the precursors and their associated fragments are therefore modulated (Figures 1a–1c). Through visual interpretation of the modulation behavior or by mathematical deconvolution, the fragments can then be assigned to the correct precursor(s).

In related work, it was previously recognized that the problems associated with chimeras can be addressed by Q modulation. A binary alternation of the Q position was implemented in a DIA LC/MS[2] method[34] and in a precursor-dependent LC/MS[2] method.[17] Similar to IQAROS, a Q movement with a small step size was reported for two DIA LC/MS[2] methods termed scanning SWATH[35] and SONAR.[36] Notably, scanning SWATH, SONAR and IQAROS have in common that they resolve chimeras by Q modulation but they also have important differences. Scanning SWATH and SONAR are DIA methods for LC/MS, where co-eluting precursors are modulated by moving a broad Q window[35,36] (m/z 10–24) over the entire mass range to fragment all eluting compounds. In contrast, IQAROS is a method for DI-MS, which modulates with the narrowest possible Q width (here, m/z 0.4) only a targeted precursor of interest and interfering isobars, e.g., a MS[1] feature which was statistically significant in a metabolomic study[4] together with neighboring signals. Also notable is the DI-HRMS study of Wang et al,[37] who visually interpreted how the signal intensities of two interfering precursors in their petroleum sample changed upon positioning the Q at two to three different locations. From this, they concluded which fragment belongs to which precursor. To the best of our knowledge, modulation with the Q isolation window's center has never been performed systematically in DI-MS. In this study, multiple precursors are modulated over numerous small Q steps followed by a mathematical deconvolution. We first assess the performance of IQAROS by analyzing mixtures of isobaric standards with ESI and then we apply IQAROS to breath analysis with secondary electrospray ionization (SESI), which is a DI metabolomic method where a key bottleneck is biomarker identification from the complex samples.[4]

2 | METHODS

2.1 | Chemicals

For preparation of the ESI and SESI buffer solution, water (H_2O, Optima, Fisher Chemical, LC/MS grade), methanol (MeOH, Optima, Fisher Chemical, LC/MS grade) and formic acid (FA, Merck, for analysis, purity 98–100%) were used. As model compounds, six readily available isobars, which are separable in MS[1] but co-fragment in MS[2], were selected and are shown in Figure 2. Namely, benzothiazole (1, TCI, purity >96.0%), pyridine-2,6-dicarbaldehyde (2, Fluorochem, purity ≥98%), 3H-pyrrolo[2,3-d]pyrimidin-4(7H)-one (3, Fluorochem, purity ≥95%), adenine (4, Sigma-Aldrich, purity ≥99%), acetanilide (5, Sigma-Aldrich, purity ≥99.5%) and N,N-dimethylbenzylamine (6, Sigma-Aldrich, purity ≥99%) were used. For an exemplary application, azelaic acid (7, Sigma-Aldrich, purity ≥98%) and 10-hydroxydecanoic acid (8, Apollo Scientific, purity ≥95%), which were previously identified in a breath metabolomic study,[38] were used for control measurements. Commercial calibration solutions (Pierce ESI Positive Ion Calibration Solution and Pierce ESI Negative Ion Calibration Solution, Thermo Scientific) were used to mass calibrate the instrument with the ESI source.

2.2 | Sample preparation

The SESI spray solution consisted of H_2O + 0.1% FA. For ESI, a 50:50 (v/v) MeOH/H_2O + 0.1% FA solution was prepared as a blank and to dissolve the standards. A 3 μM solution was prepared for standards 1 and 3 to 6 and a 6 μM solution for standards 2, 7 and 8 to analyze them individually. Accounting for the experimentally determined sensitivities of the standards 1 to 6, mixtures with equal signal
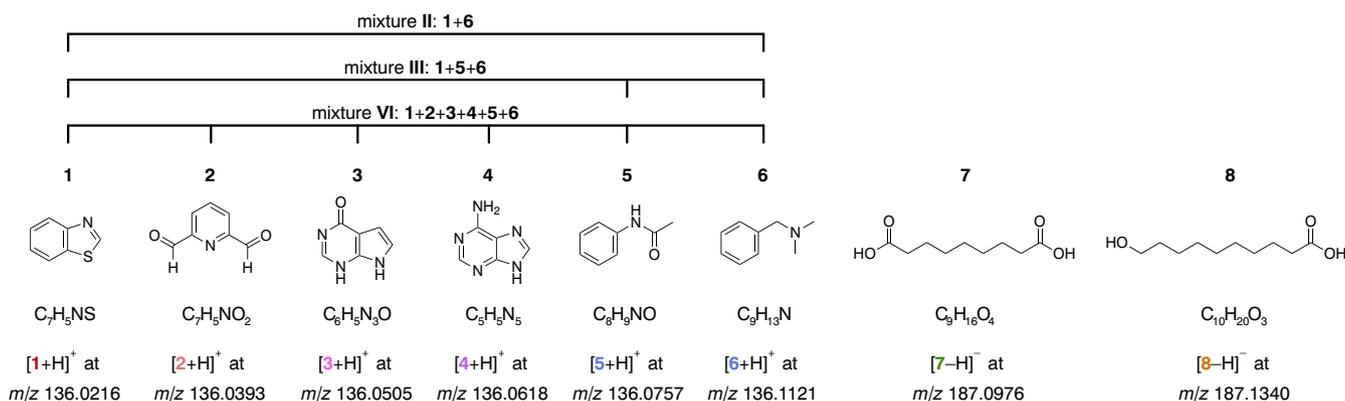
FIGURE 2 Compounds used in this study. Their labels, structures, molecular formulae and exact monoisotopic masses for a given ionization mode are listed [Color figure can be viewed at wileyonlinelibrary.com]

intensities were prepared: a two isobars mixture called mixture **II** consisting of 9.82 μM **1** and 0.35 μM **6**; a three isobars mixture called mixture **III** consisting of 9.82 μM **1**, 0.77 μM **5** and 0.35 μM **6**; and a six isobars mixture called mixture **VI** consisting of 9.82 μM **1**, 14.80 μM **2**, 2.51 μM **3**, 0.97 μM **4**, 0.77 μM **5** and 0.35 μM **6**.

2.3 | Mass spectrometry

The mass spectrometer used in this study was an orbitrap Q-Exactive Plus (Thermo Scientific) operated with the manufacturer's standard control software (ExactiveTune, version 2.9, Thermo Scientific) and Xcalibur (version 4.1.31.9, Thermo Scientific). Mass calibration was performed according to the instrument manual and was always more recent than 7 days according to specifications.

MS^1 spectra were acquired with the following settings: narrowest scan range with a m/z 0.4 isolation width around the target center, profile mode, 5e6 automatic gain control (AGC) target, 500 ms maximum injection time (IT). Where not otherwise specified, the lowest resolution setting at 17,500 was selected.

Direct MS^2 spectra were acquired with the following settings: narrowest isolation width m/z 0.4 centered around the monoisotopic exact mass of the investigated compounds, profile mode, 3e6 AGC target, 500 ms maximum IT, 10, 35 and 50 eV stepped collision energies (CE) and 17,500 resolution.

Similar settings were used for IQAROS, but instead of centering around the monoisotopic exact mass, incremental target masses were manually entered via the graphical user interface of Xcalibur (Figure S1, supporting information). The list started −0.6 Da and ended +0.6 Da of the nominal mass of interest and contained entries in steps of 0.02 Da. For example, to apply IQAROS to an ion at m/z 136.11207, the list was constructed as follows: 135.40, 135.42, 135.44, …, 136.56, 136.58, 136.60 and contained a total of 61 entries. Moreover, a 500 ms maximum IT was used. The AGC target was set to 3e6 except for mixture **VI**, where it was set to 5e5 due to some intensity regulation artefact at AGC target 3e6. For the lower abundant SESI biomarkers, a repetition of every step (122 entries instead of 61) was used.

ESI measurements were performed with a standard source for this instrument (HESI-II probe, Thermo Scientific). The ESI settings were: (+)- or (−)-polarity, 4 kV spray voltage, 320°C ion transfer capillary temperature, 12 psi sheath gas, 0 aux gas, 0 sweep gas and 50 S-lens RF level. Breath analysis with SESI was performed with a commercial SESI source (Super SESI, Fossiliontech) equipped with a 20 μm i.d. capillary (for Super SESI, Fossiliontech) and a flow sensor (Exhalion, Fossiliontech). Similar to prior studies with the same setup,[10,11] a subject exhaled through a spirometry filter (MicroGard IIB, Vyaire Medical) with 8 L min^{-1} of which 0.3 L min^{-1} was directed through a transfer line at 130°C into an ionization chamber held at 90°C. For SESI, the analyzer's settings were: (−)-polarity, 3.5 kV spray voltage, 250°C ion transfer capillary temperature and 60 S-lens RF level.

2.4 | Data analysis

MS^1.raw files were converted into.mzXML files with MSConvert (version 3, ProteoWizard[39]), processed and plotted in Matlab (R2018a, Mathworks). In contrast, MS^2 .raw files were converted into.mgf file format with MSConvert and processed with Matlab.

Direct ESI-MS^2 data for standards and the blank were averaged over all scans for each selected precursor and peak-picked with the Matlab function mspeaks. The MS^2 spectra of the standards were then blank subtracted. The resulting averaged and blank-subtracted MS^2 spectra were written into a new.mgf file for every precursor and used to compare the results against the IQAROS output. The direct SESI-MS^2 was averaged during exhalation and baseline-subtracted. Then, the same processing as for the ESI standards was applied.

An overview of the IQAROS code is shown in Figure S2 (supporting information). Moreover, the Matlab code (provided as part of the archived data under DOI 10.3929/ethz-b-000520528) is fully commented; thus, only a brief description is given here. In a first step, the code determines which peaks are located within the m/z range where Q modulation took place. These peaks are considered precursors. Peaks with a lower m/z than the precursors are considered fragments. For all fragments and for all intense precursors (>5% of the maximum precursor intensity), the extracted ion currents (XICs) are calculated. For p scans and n fragments, the XICs are written into a $n \times p$ matrix herein called **F**. Likewise, the XICs for m precursors are written into a $m \times p$ matrix called **P**.

Similar to the work of Nikolskiy et al,[17] the underlying model is that the intensity $F_{i,s}$ of the fragment i in scan s can be expressed as a linear combination of all m precursor intensities $P_{j,s}$ in the same scan s multiplied with a contribution coefficient $\beta_{i,j}$, i.e., $F_{i,s} = \sum_{j=1}^{m} \beta_{i,j} \cdot P_{j,s}$. As an example, Figures 1c and 1d depict $n = 3$ fragments, $m = 2$ precursors and $p = 3$ scans. Instead of considering only one scan s and one fragment i, this can be rewritten for all p scans and n fragments: $\mathbf{F} = \boldsymbol{\beta} \cdot \mathbf{P}$ with the $n \times p$ matrix **F** containing the fragment intensities for n fragments and p scans, the $m \times p$ matrix **P** containing the precursors intensities for m precursors and p scans, and the $n \times m$ matrix $\boldsymbol{\beta}$ describing the contribution to fragment i from every precursor j. The column j of matrix $\boldsymbol{\beta}$ represents the MS^2 spectrum for precursor j. For example, $\beta_{i,j} = 0$ means that fragment i is not part of the MS^2 spectrum of precursor j. Likewise, $\beta_{i,j} = 1$ or $\beta_{i,j} = 2$ signifies that fragment i appears in the MS^2 spectrum of precursor j with the same or double the intensity of precursor j, respectively.

To obtain the estimates $\widehat{\beta}_{i,j}$, two methods were used, as depicted in Figure 1d: method 1 is identical to that of Nikolskiy et al,[17] i.e., $\widehat{\beta}_{i,j}$ are estimated by running a non-negative multiple linear regression with the Matlab function lsqnonneg for every fragment i. Performing n times a non-negative m-multiple linear regression yields matrix $\hat{\boldsymbol{\beta}}$. For method 2, it is assumed that every fragment is the product of only one precursor. The aim is to find the single precursor j which alone best describes the observation of fragment i. Thus, for every ith fragment m, non-negative simple linear regressions are run with every jth precursor XIC as independent variable yielding in m $\widehat{\beta}_{i,j}$ per

fragment i. For every jth regression, the coefficient of determination $R^2_{i,j}$ is calculated. Only the regression coefficient $\hat{\beta}_{i,j}$ with the highest $R^2_{i,j}$ is written into matrix $\hat{\beta}$ for fragment i, all other values of $\hat{\beta}_{i,j}$ are set to zero. Thus, every row i of $\hat{\beta}$ contains only one element >0.

From the $\hat{\beta}$ matrix, the mass spectra of the precursors are reconstructed, i.e. for a column j in $\hat{\beta}$ the entry $\hat{\beta}_{i,j}$ represents the intensity of fragment i when precursor j has a normalized intensity of 1. Finally, the reconstructed mass spectra are saved as .mgf files.

2.5 | Spectral matching

The deconvoluted spectra were compared with the blank-subtracted direct MS^2 spectra. For this purpose, peaks with higher m/z values than the precursor were omitted and the intensities were filled into two incremental vectors $\vec{I}_1$ and $\vec{I}_2$ with 0.005 m/z increments. The MS^2 match score was then calculated as the squared cosine of the angle θ between the two vectors[40,41]:

$$\text{score} = \cos^2(\theta) = \left(\frac{\vec{I}_1 \cdot \vec{I}_2}{\left\| \vec{I}_1 \right\| \left\| \vec{I}_2 \right\|} \right)^2.$$

To further assess the performance of IQAROS, the direct and the deconvoluted MS^2 spectra were processed with SIRIUS[24,25] (version 4.8.2). The database search was performed in all available databases for $[M + H]^+$ or $[M - H]^-$ species, respectively. Otherwise, orbitrap default settings were used except for the MS^2 mass tolerance (MS2 MassDev) which was set to 10 ppm and proposed structures with electron sextets were ignored.

3 | RESULTS AND DISCUSSION

3.1 | Deconvolution performance with isobaric standards

The problem of chimeric MS^2 spectra is shown in Figure 3a. The exemplary isobars **1** and **6** in mixture **II** are perfectly separable on the MS^1 level. When performing classical direct MS^2, i.e. centering the quadrupole at the precursor of interest, here m/z 136.0216 for **1** and m/z 136.1121 for **6**, problems with chimeras arise. Since the two ions only differ by m/z 0.09, they cannot be isolated separately by the Q – even at the narrowest isolation width of m/z 0.4. In fact, when targeting **1** or **6** with direct MS^2, the two spectra are contaminated by
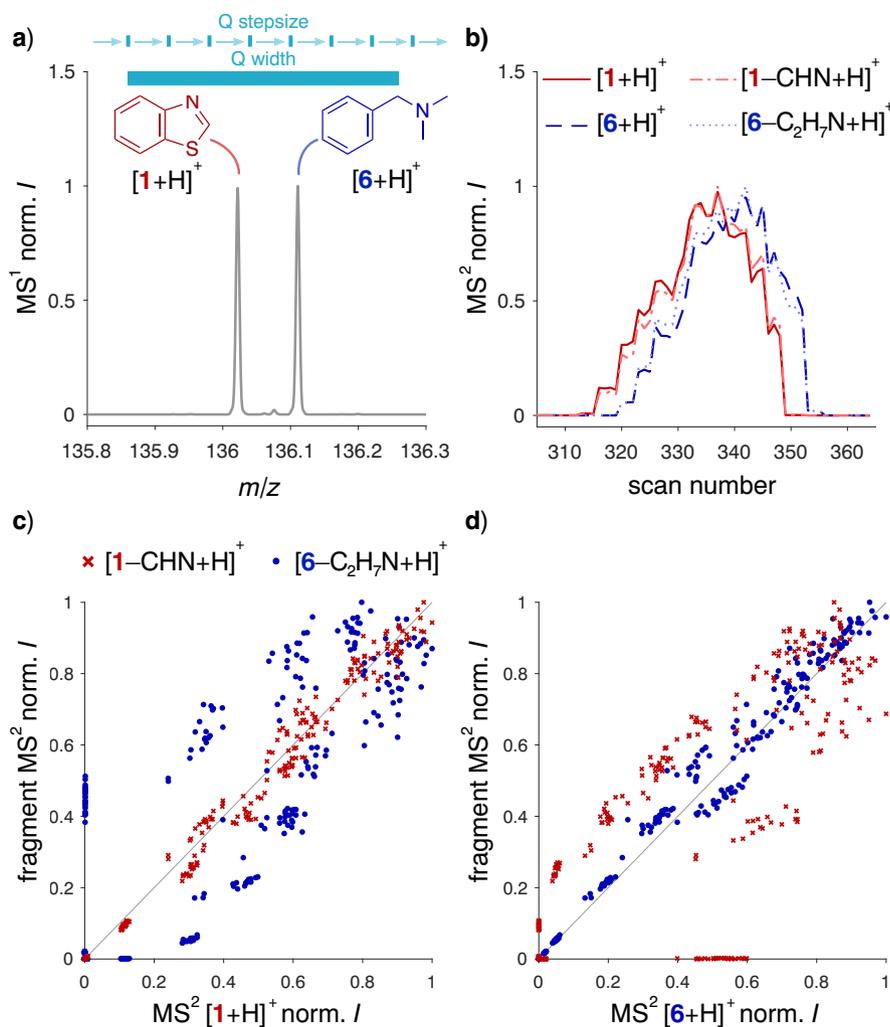
FIGURE 3 Overview of IQAROS applied to mixture **II** containing **1** and **6**. a, The MS^1 signal of the isobar mixture (the m/z 0.4 wide Q isolation window as well as the m/z 0.06 Q step size accuracy are also depicted). b, The normalized MS^2 XICs of $[1 + H]^+$ and $[6 + H]^+$ as well as their major fragments $[1 - CHN + H]^+$ and $[6 - C_2H_7N + H]^+$ during one modulation cycle with the Q isolation window. The XICs over all modulation cycles are shown in Figure S3a (supporting information). c, d, Plots of the normalized fragment intensities against the normalized intensities of the precursor $[1 + H]^+$ and $[6 + H]^+$, respectively. The three-dimensional representation of the same plot can be found in Figure S3b (supporting information) [Color figure can be viewed at wileyonlinelibrary.com]

the other isobar. Particularly, the spectrum of **1** is severely contaminated by the tropylium ion $[6 - C_2H_7N + H]^+$ originating from **6** (Figure S4a, supporting information).

While m/z 0.4 is the narrowest Q isolation width, the Q isolation center can be entered in the instrument control software with a theoretical accuracy of m/z 0.00001. However, we found on our instrument that the Q hardware reacts to input by the software with an accuracy of approximately m/z 0.06, although this does not limit the resolving power of IQAROS, as will be discussed below. Thus, we set up IQAROS with a software-defined arbitrary increment of m/z 0.02 (Figure S1, supporting information). Important is that the software-defined increment is smaller than the hardware increment; here 0.02 ≤0.06, but different values are expected on other instruments. A larger step size within the hardware limit could make IQAROS even faster, but this was not optimized here. Therefore, the actual modulation corresponds to the one depicted in Figure 1b, i.e., the software moves the Q window after every scan by m/z 0.02 (Figure S1, supporting information) but the hardware reacts only after every third scan by moving m/z 0.06 ahead. In principle, an alternation between MS^1 and MS^2 is also possible, which would result in a method as depicted in Figure S5a (supporting information) and would be better to detect low-abundance interfering isobars in the MS^1 scan. However, due to an instrument-specific technical restriction, which does not allow the orbitrap scan range and the Q isolation range for MS^1 to be set independently from each another (details in Figure S5b and S5c,

supporting information), the modulation here was only performed on the MS^2 level, as shown in Figure 1b.

Figure 3b shows the XICs of $[1 + H]^+$ and $[6 + H]^+$ in mixture **II** as well as their major fragments $[1 - CHN + H]^+$ and $[6 - C_2H_7N + H]^+$ over one modulation cycle. The total modulation involved six of these cycles and required approximately 360 scans or 3 min in total (Figure S3a, supporting information). The precursor-fragment correlation can be depicted by plotting the normalized intensities of the fragments against the normalized intensities of the precursors as shown in Figures 3c and 3d or in three dimensions in Figure S3b (supporting information). $[1 - CHN + H]^+$ correlates well with $[1 + H]^+$ and also with $[6 + H]^+$, but with significantly higher residuals for the latter. Thus, a relationship of $[1 - CHN + H]^+$ with **1** but not with **6** can be deduced. Similar observations can be made with mixture **III** with **1**, **5** and **6** (Figure S6, supporting information) and mixture **VI** with **1**, **2**, **3**, **4**, **5** and **6** (Figure S7, supporting information). Remarkably, even if isobars fall into the same m/z 0.06 Q step size, they and their fragments can be distinguished because their intensities are modulated distinctively, e.g., in scan 473 the normalized intensity ratios of $[3 + H]^+$, $[4 + H]^+$ and $[5 + H]^+$ are 3%:10%:20% (Figure S7c, supporting information), although they are only spaced by m/z 0.025. Consequently, IQAROS has the potential to deconvolute isobars even if they are spaced closer to each other than the accuracy with which the Q's center of mass can be set.

To extract this relationship quantitatively, the two deconvolution methods were applied: method 1 establishes a fragment-precursor
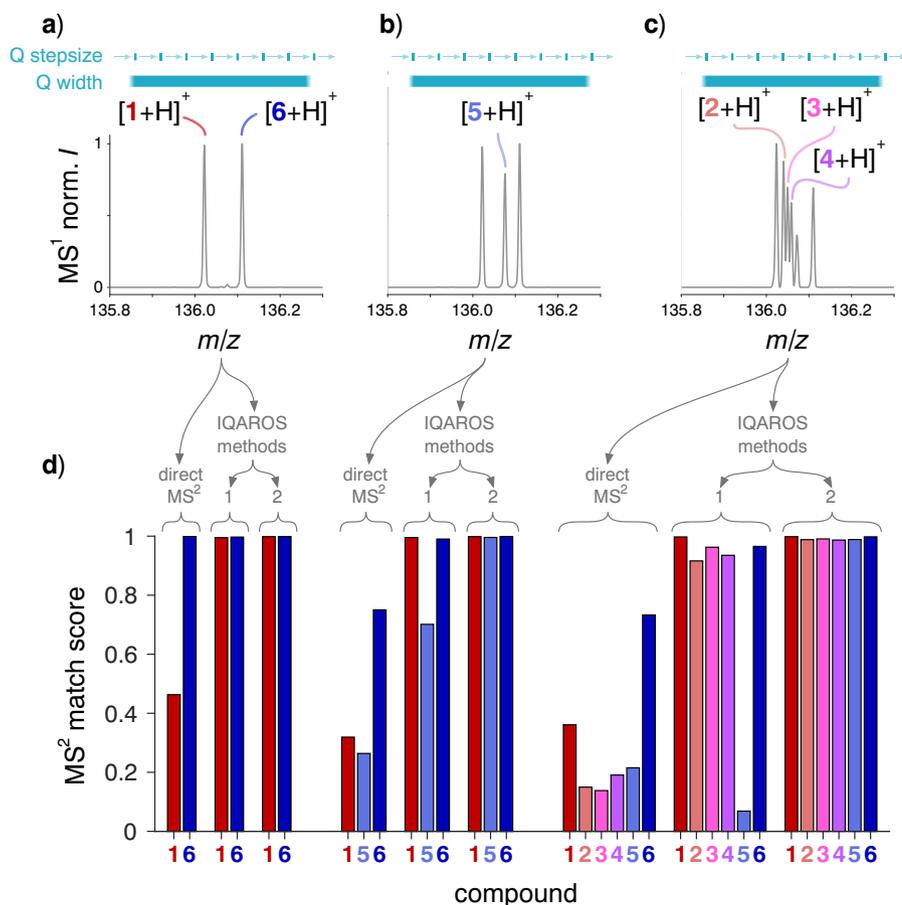
FIGURE 4 MS^2 match scores of the direct and IQAROS deconvoluted spectra for mixtures **II**, **III** and **VI**. a, b, and c, MS^1 spectra for mixtures **II**, **III** and **VI**, respectively. d, MS^2 match scores between MS^2 spectra of the pure standards with the direct and the IQAROS spectra of the isobaric mixtures. For IQAROS, results for deconvolution methods 1 and 2 are shown [Color figure can be viewed at wileyonlinelibrary.com]

relationship by describing every fragment intensity as a positive linear combination of precursor intensities. In contrast, method 2 assigns every fragment to that single precursor which best explains the intensity of the fragment. Mathematically, method 1 is based on a non-negative multi-linear regression. The reconstructed spectra of mixtures **II**, **III** and **VI** are compared with the pure standards in Figure S4, S8 and S9 (supporting information), respectively. For benchmarking with the standard method, the pure standards are also compared with the direct MS^2 measurements of the isobaric mixtures. The direct and the IQAROS spectra are compared with the pure standards by calculating the MS^2 match score, which is plotted in Figure 4. Except for compound **5** in mixture **VI**, all constellations have a higher MS^2 match score with IQAROS with deconvolution method 1. Indeed, the mean MS^2 match score for all eleven constellations for direct MS^2 is 0.42 and for IQAROS MS^2 0.87. Interestingly, the score for **5** in mixture **III** is better than in mixture **VI**. In fact, with mixture **VI**, the deconvolution algorithm incorrectly assigns the fragment $[5 - C_2H_2O + H]^+$ to **3**, **4** and **6** instead of **5**. Hence, a non-negative linear combination of the three precursors seems to describe the XIC $[5 - C_2H_2O + H]^+$ better than **5** alone. This finding can be interpreted as a problem with multicollinearity, i.e., the independent variables themselves are correlated with each another. This is the case for IQAROS where moving Q's isolation center into the isobars region increases the intensity of all precursors together. Multicollinearity can be quantified with a mathematical figure of merit called the variance inflation factor, $0 \leq VIF < \infty$, where a value of VIF >10 is considered problematic in regards of multicollinearity.[42] For mixture **II**, the two VIFs are around 10^{-5}. For mixture **III**, the VIFs of **1** and **6** are at 7 and 11 and for **5** already at 120 indicating problematic multicollinearity.

For mixture **VI**, the VIFs lie between 16 and 1060. Consequently, the poor result for **5** in mixture **VI** can be attributed to issues with multicollinearity of the method and the deconvolution model.

Multicollinearity can be tackled by alternative regression methods like ridge or lasso (least absolute shrinkage and selection operator).[43,44] However, these methods cannot be run on the standard numerical computing software used here with the non-negativity restriction. Yet, determining $\hat{\beta}$ in this study is restricted to non-negativity because it is assumed that a precursor never leads to a "negative intensity" fragment. This assumption is important, because it already eliminates some mathematical linear combinations which would otherwise be troublesome when dealing with the multicollinear data. Thus, an alternative to ridge or lasso regressions when dealing with multicollinearity is variable elimination in the model.[43] This is implemented in method 2, where every fragment is assigned to the single precursor which best explains the XIC of the given fragment. Mathematically, this is done by picking the non-negative simple linear regression model with the best fit, i.e., with the highest coefficient of determination. Method 2 will not suffer from multicollinearity as method 1. However, as a disadvantage, method 2 will not correctly describe when two precursors lead to the same fragment. Instead, the fragment will be assigned to only one precursor as, for example, depicted for fragment F_3 in Figures 1c and 1d. The IQAROS spectra deconvoluted by method 2 are shown for the three isobaric mixtures in Figures S10, S11 and S12 (supporting information) and the MS^2 match scores comparison in Figure 4. With method 2, IQAROS yields an average MS^2 match score of 0.99 and therefore the multicollinearity and the problem with compound **5** in mixture **VI** is eliminated.

TABLE 1 Compound identification of the standards using SIRIUS.[24,25] The first column shows the structures of the six isobaric standards. The second column shows the most likely structures according to the output of SIRIUS when the pure blank-subtracted standards are processed. If the true structure is not the top hit, the position of the correct hit is listed instead. The third column shows the SIRIUS output when processing the direct MS^2 targeted mixture **VI**. The fourth and fifth columns show the SIRIUS output of mixture **VI** when analyzed by IQAROS with method 1 and method 2, respectively

used standards			pure standards with direct MS^2			mixture **VI** with direct MS^2			mixture **VI** with IQAROS method 1			mixture **VI** with IQAROS method 2		
label	molecular formula	structure	molecular formula	structure	correct hit position	molecular formula	structure	correct hit position	molecular formula	structure	correct hit position	molecular formula	structure	correct hit position
1	C_7H_5NS		C_7H_5NS		1	C_7H_5NS		1	C_7H_5NS		1	C_7H_5NS		1
2	$C_7H_5NO_2$		$C_7H_5NO_2$		7	$C_7H_5NO_2$		13	$C_7H_5NO_2$		8	$C_7H_5NO_2$		7
3	$C_6H_5N_3O$		$C_6H_5N_3O$		18	$C_6H_5N_3O$		18	$C_6H_5N_3O$		20	$C_6H_5N_3O$		81
4	$C_5H_5N_5$		$C_5H_5N_5$		1	$C_4H_9NO_4$			$C_5H_5N_5$		1	$C_5H_5N_5$		1
5	C_8H_9NO		C_8H_9NO		1	$C_5H_{11}O_4$	only structures noncompliant with electron octet rule		C_8H_9NO		not in top 100	C_8H_9NO		3
6	$C_9H_{13}N$		$C_9H_{13}N$		1	$C_9H_{13}N$		1	$C_9H_{13}N$		1	$C_9H_{13}N$	no search performed	

3.2 | Spectral search with isobaric standards

In the previous section, it was demonstrated that IQAROS leads to fragment spectra comparable with spectra of the pure standards. Here, the discussion focuses on how helpful this is for compound identification. For this purpose, the previously discussed spectra were processed with the spectral interpretation software SIRIUS.[24,25] The most challenging sample, i.e., mixture **VI**, is addressed.

The first column in Table 1 shows the structures of the six isobaric standards. When the pure standards are analyzed by direct MS^2, blank-subtracted and processed with SIRIUS, the structures shown in the second column in Table 1 are obtained. This represents the best possible outcome because the inputs are the pure standards. In fact, SIRIUS correctly assigns 6 of 6 molecular formulae and 4 of 6 structures. The two incorrectly assigned structures are closely related isomers of the true structures. In contrast, when mixture **VI**, analyzed with direct MS^2, is processed, only 4 of 6 molecular formulae and 2 of 6 structures are correct. The wrong assignment of the molecular formulae arises from a combination of two adverse effects in high-resolution tandem MS: (1) partial ion coalescence[45-47]

leads to significant shifts in accurate mass of the precursor (Figure S13, supporting information) and (2) the many fragments from other precursors possibly coincided with a fragment which would be consistent with the erroneous precursor mass. Here, IQAROS can help by eliminating fragments which coincidently fit a reasonable mass difference but actually stem from an interfering precursor. Indeed, 6 of 6 molecular formulae are assigned correctly with method 1 IQAROS. Moreover, 3 of 6 structures are correct – comparable with the 4 of 6 when the pure standards are measured. Only **5** is not reasonably assigned. Presumably, the aforementioned elimination of $[5 - C_2H_2O + H]^+$ from the fragment list of **5** due to multicollinearity renders the structural assignment difficult because $[5 - C_2H_2O + H]^+$ would indicate a N-monosubstituted phenyl ring (N analog of the tropylium ion).

If IQAROS with method 2 is applied, the true structure of **5** is at the 3rd hit position. As a disadvantage, method 2 eliminates too many features for **6**. While it correctly assigns the main fragment $[6 - C_2H_7N + H]^+$ to **6**, it omits the minor fragment $[6 - C_4H_9N + H]^+$, which corresponds to $[C_5H_4 + H]^+$ and is found for many aromatic compounds. Only with one fragment was SIRIUS not able to
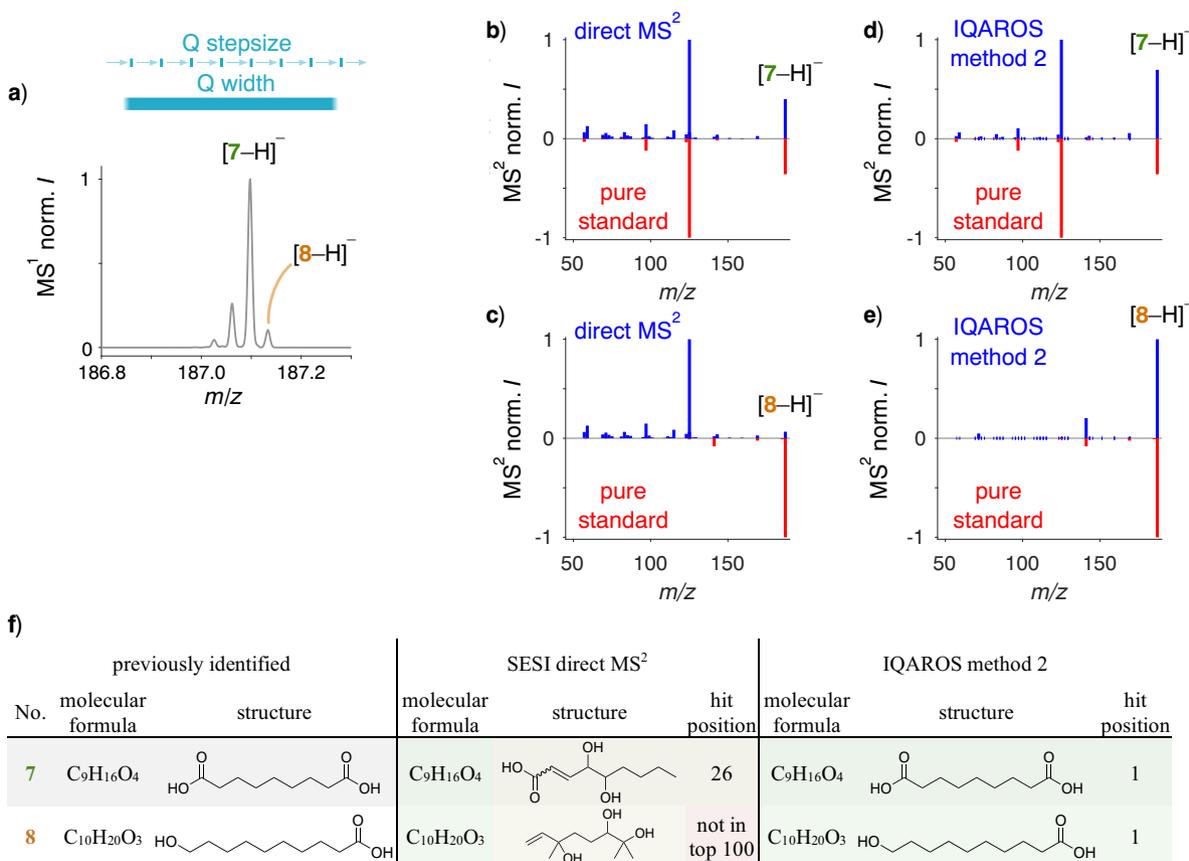
FIGURE 5 Example of two isobaric biomarkers in breath analyzed by SESI-MS and identified with IQAROS: a, SESI-MS^1 range during exhalation around m/z 187 where the previously identified biomarkers **7** and **8** lay.[38] b, c, Direct SESI-MS^2 spectra of **7** and **8** in blue. For comparison, the ESI-MS^2 of the pure standards is shown in red. d, e, SESI spectra obtained with IQAROS method 2 for **7** and **8** in blue and the same ESI-MS^2 spectra as before in red. F, In the first column, the previously identified biomarkers[38] are shown, in the second column, the SIRIUS output from the direct MS^2 spectra b) and c) is shown, and, in the third column, the SIRIUS output from IQAROS from d) and e) is depicted [Color figure can be viewed at wileyonlinelibrary.com]

run a search and thus left **6** unassigned. In summary, IQAROS with method 2 can avoid errors from multicollinearity but meanwhile might be too conservative in fragment elimination.

In conclusion, both IQAROS methods 1 and 2 lead to better formula assignments when analyzing isobaric mixtures compared with the classical direct MS^2 analysis of the mixed isobars. For structure assignment of the mixed isobars, IQAROS performs in most cases better or comparable to direct MS^2 analysis. However, the structure assignment does not meet the quality of the pure standards analyzed individually by direct MS^2.

3.3 | Example: Identification of isobars in breath

An exemplary application will now be discussed where IQAROS turns out to be very useful. Secondary electrospray ionization (SESI)-MS is an enormously sensitive and rapid technique for real-time DI analysis of volatile metabolites and therefore is typically applied for breath analysis.[4] Depending on the acquisition technique and instrument, a few hundred to a few thousand features are found in the mass range from *m/z* 50 to 500 during exhalation.[10,11] Taking into account that there are also some background peaks, it becomes apparent that many features must have neighboring isobars and thus often account for ≤50% of the isolated total intensity, which makes MS^2 spectral matching ambiguous.[9] Hence, reliable biomarker identification is usually done by collecting breath condensate followed by LC/ESI-MS.[38] However, this method is time-consuming and can lead to contamination/analyte loss during sample preparation. As a consequence, it is highly desirable to obtain higher quality MS^2 spectra of biomarkers without the need for condensation and LC separation.

As an example, isobaric signals around *m/z* 187 acquired by SESI-MS during the exhalation of a volunteer are shown in Figure 5a. In a previous study by Gaugg et al,[38] two of the isobars were identified by breath condensation and LC/MS to correspond to fatty acids **7** and **8**. Both in the study of Gaugg et al[38] as well as here (Figures 5b and 5c in blue), a direct MS^2 experiment acquired over five exhalations leads to non-interpretable fragment spectra due to the co-fragmented isobars. In fact, if the direct MS^2 spectra are processed by SIRIUS, incorrect structural assignments are obtained (Figure 5f, second column). Conversely, with IQAROS over five exhalations and the more stringent method 2, the spectra (Figures 5d and 5e in blue) fit well with the ESI-MS^2 spectra obtained from the pure standards (Figures 5d and 5e) in red). If the IQAROS spectra are processed with SIRIUS, they yield the same structures as the ones reported by Gaugg et al.[38]

In conclusion, direct SESI-MS^2 can lead to incorrect assignments due to co-fragmented isobars. In contrast, with IQAROS biomarkers in SESI-MS can be accurately identified without the need to collect breath condensate followed by LC/MS. Comparable results are obtained within minutes instead of several hours of sample preparation and analysis.

4 | CONCLUSIONS

IQAROS (incremental quadrupole acquisition to resolve overlapping spectra) is an approach to disentangle chimeric MS^2 spectra that arise when a precursor of interest is co-fragmented with neighboring isobars in DI-HRMS. The method modulates precursor and fragment intensities by moving the isolation quadrupole's center over the precursor range in a stepwise fashion, followed by a mathematical deconvolution to reconstruct the MS^2 spectra of the individual precursors. The method can easily be implemented by simply creating a list with incremental precursor isolation center masses using the graphical user interface of the instrument.

For the analysis of a mixture containing up to six isobaric standards confined within a range of *m/z* 0.09, IQAROS delivered higher quality MS^2 data compared with a direct MS^2 approach. The match between the deconvoluted spectra and the spectra of pure standards was significantly higher and compound search with the spectra interpretation software SIRIUS led to a better hit rate. Depending on the mathematical model, problems might arise from multicollinearity (method 1 – multi linear regression) or from the incapability of the deconvolution algorithm to assign a fragment to more than one precursor (method 2 – simple linear regression). As an exemplary application, IQAROS was able to directly identify two biomarkers from on-line breath measurements yielding the same structures as in a previous study,[38] but without the need for time-consuming breath condensate collection and LC/MS analysis. In general, IQAROS might be useful for all sorts of ambient or desorption ionization methods like DART, DBDI, DESI, EESI, DAPCI, and many more.[48] For these techniques, hyphenation with chromatography to resolve interferences is often difficult or impossible. In contrast, IQAROS allows to mitigate the problem of co-fragmentation without the need for chromatography or changes in the ionization process.

In this study, an isolation quadrupole in tandem with an orbitrap was used together with ESI-based ionization sources. The deconvolution was performed by a non-negative multi- or simple-linear regression. It is important to note that IQAROS can be used with other setups as well: (1) it can be used for variable fragmentation techniques or ionization methods; (2) it can also be employed for other high-resolution MS instruments such as time-of-flight or Fourier transform ion cyclotron resonance analyzers, and (3) other mathematical methods such as ridge, lasso or many more could possibly be used for deconvolution.

Limitations arise from multicollinearity when many precursors are deconvoluted. This problem can possibly be tackled with more advanced statistical methods. Moreover, the method needs sufficiently intense precursors, i.e., if a precursor is modulated to 10% of its maximum intensity, a fragment must still be abundant enough to be detected. An additional limitation, unique to the instrument used in this study, is that at 10 V collision energy (CE) the precursors must still be detectable. However, this criterion is repealed if modulation can be performed while alternating between MS^1 and MS^2.

ACKNOWLEDGMENTS

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/rcm.9266.

DATA AVAILABILITY STATEMENT

The original data used in this publication are made available in a curated data archive at ETH Zürich (https://www.research-collection.ethz.ch) under the DOI https://doi.org/10.3929/ethz-b-000520528. There, the fully commented IQAROS deconvolution Matlab code is provided. Additionally, the modulation experiments of mixtures **II** and **III** are provided as exemplary.mgf files such that the user can familiarize with the processing.

ORCID

Renato Zenobi 🔘 https://orcid.org/0000-0001-5211-4358

REFERENCES

1. de Raad M, Fischer CR, Northen TR. High-throughput platforms for metabolomics. *Curr Opin Chem Biol*. 2016;30:7-13. doi:10.1016/j.cbpa.2015.10.012
2. Urban PL, Chen Y-C, Wang Y-S. *Time-Resolved Mass Spectrometry: From Concept to Applications*. Chichester, West Sussex: John Wiley & Sons; 2016 10.1002/9781118887332.
3. González-Domínguez R, Sayago A, Fernández-Recamales Á. Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis*. 2017;9(1):131-148. doi:10.4155/bio-2016-0202
4. Bruderer T, Gaisl T, Gaugg MT, et al. On-line analysis of exhaled breath. *Chem Rev*. 2019;119(19):10803-10828. doi:10.1021/acs.chemrev.9b00005
5. Lin L, Yu Q, Yan X, et al. Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. *Analyst*. 2010;135(11):2970-2978. doi:10.1039/C0AN00265H
6. Gaugg MT, Gomez DG, Barrios-Collado C, et al. Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry A pilot study on tobacco smokers. *J Breath Res*. 2016;10(1):016010 doi:10.1088/1752-7155/10/1/016010
7. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ Sci Technol*. 2014;48(4):2097-2098. doi:10.1021/es5002105
8. Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, Old WM. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. *J Proteome Res*. 2010;9(8):4152-4160. doi:10.1021/pr1003856
9. Lawson TN, Weber RJ, Jones MR, et al. msPurity: Automated evaluation of precursor ion purity for mass spectrometry-based fragmentation in metabolomics. *Anal Chem*. 2017;89(4):2432-2439. doi:10.1021/acs.analchem.6b04358
10. Singh KD, Tancev G, Decrue F, et al. Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry. *Anal Bioanal Chem*. 2019;411(19):4883-4898. doi:10.1007/s00216-019-01764-8
11. Lan J, Kaeslin J, Greter G, Zenobi R. Minimizing ion competition boosts volatile metabolome coverage by secondary electrospray ionization orbitrap mass spectrometry. *Anal Chim Acta*. 2021;1150:338209 doi:10.1016/j.aca.2021.338209
12. Colby BN. Spectral deconvolution for overlapping GC/MS components. *J Am Soc Mass Spectrom*. 1992;3(5):558-562. doi:10.1016/1044-0305(92)85033-G
13. Tsou C-C, Avtonomov D, Larsen B, et al. DIA-umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods*. 2015;12(3):258-264. doi:10.1038/nmeth.3255
14. Samanipour S, Reid MJ, Bæk K, Thomas KV. Combining a deconvolution and a universal library search algorithm for the nontarget analysis of data-independent acquisition mode liquid chromatography−high-resolution mass spectrometry results. *Environ Sci Technol*. 2018;52(8):4694-4701. doi:10.1021/acs.est.8b00259
15. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;12(6):523-526. doi:10.1038/nmeth.3393
16. Li H, Cai Y, Guo Y, Chen F, Zhu Z-J. MetDIA: Targeted metabolite extraction of multiplexed MS/MS spectra generated by data-independent acquisition. *Anal Chem*. 2016;88(17):8757-8764. doi:10.1021/acs.analchem.6b02122
17. Nikolskiy I, Mahieu NG, Chen Y-J, Tautenhahn R, Patti GJ. An untargeted metabolomic workflow to improve structural characterization of metabolites. *Anal Chem*. 2013;85(16):7713-7719. doi:10.1021/ac400751j
18. Dorfer V, Maltsev S, Winkler S, Mechtler K. CharmeRT: Boosting peptide identifications by chimeric spectra identification and retention time prediction. *J Proteome Res*. 2018;17(8):2581-2589. doi:10.1021/acs.jproteome.7b00836
19. Peckner R, Myers SA, Jacome ASV, et al. Specter: Linear deconvolution for targeted analysis of data-independent acquisition mass spectrometry proteomics. *Nat Methods*. 2018;15(5):371-378. doi:10.1038/nmeth.4643
20. Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat Methods*. 2021;18(7):779-787. doi:10.1038/s41592-021-01195-3
21. Allen F, Greiner R, Wishart D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics*. 2015;11(1):98-110. doi:10.1007/s11306-014-0676-4
22. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform*. 2010;11(1):148 doi:10.1186/1471-2105-11-148
23. da Silva RR, Dorrestein PC, Quinn RA. Illuminating the dark matter in metabolomics. *Proc Natl Acad Sci*. 2015;112(41):12549-12550. doi:10.1073/pnas.1516878112
24. Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods*. 2019;16(4):299-302. doi:10.1038/s41592-019-0344-8
25. 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*. 2015;112(41):12580-12585. doi:10.1073/pnas.1509788112
26. Schmid R, Petras D, Nothias L-F, et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS

environment. *Nat Commun*. 2021;12(1):3832 doi:10.1038/s41467-021-23953-9

27. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem*. 2012;84(1):283-289. doi:10.1021/ac202450g

28. Mahieu NG, Patti GJ. Systems-level annotation of a metabolomics data set reduces 25 000 features to fewer than 1000 unique metabolites. *Anal Chem*. 2017;89(19):10397-10406. doi:10.1021/acs.analchem.7b02380

29. Nash WJ, Dunn WB. From mass to metabolite in human untargeted metabolomics: Recent advances in annotation of metabolites applying liquid chromatography-mass spectrometry data. *Trends Anal Chem*. 2019;120:115324 doi:10.1016/j.trac.2018.11.022

30. Sindelar M, Patti GJ. Chemical discovery in the era of metabolomics. *J Am Chem Soc*. 2020;142(20):9097-9105. doi:10.1021/jacs.9b13198

31. Dwivedi P, Wu P, Klopsch SJ, Puzon GJ, Xun L, Hill HH. Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics*. 2008;4(1):63-80. doi:10.1007/s11306-007-0093-z

32. Jeanne Dit Fouque D, Maroto A, Memboeuf A. Purification and quantification of an isomeric compound in a mixture by collisional excitation in multistage mass spectrometry experiments. *Anal Chem*. 2016;88(22):10821-10825. doi:10.1021/acs.analchem.6b03490

33. Jeanne Dit Fouque D, Maroto A, Memboeuf A. Structural analysis of a compound despite the presence of an isobaric interference by using in-source collision induced dissociation and tandem mass spectrometry. *J Mass Spectrom*. 2021;56(2):e4698 doi:10.1002/jms.4698

34. Amodei D, Egertson J, MacLean BX, et al. Improving precursor selectivity in data-independent acquisition using overlapping windows. *J Am Soc Mass Spectrom*. 2019;30(4):669-684. doi:10.1007/s13361-018-2122-8

35. Messner CB, Demichev V, Bloomfield N, et al. Ultra-fast proteomics with scanning SWATH. *Nat Biotechnol*. 2021;39(7):846-854. doi:10.1038/s41587-021-00860-4

36. Moseley MA, Hughes CJ, Juvvadi PR, et al. Scanning quadrupole data-independent acquisition, part a: Qualitative and quantitative characterization. *J Proteome Res*. 2018;17(2):770-779. doi:10.1021/acs.jproteome.7b00464

37. Wang W, Dong M, Song C, et al. Structural information of asphaltenes derived from petroleum vacuum residue and its hydrotreated product obtained by FT-ICR mass spectrometry with narrow ion isolation windows. *Fuel*. 2018;227:111-117. doi:10.1016/j.fuel.2018.04.064

38. Gaugg MT, Bruderer T, Nowak N, et al. Mass-spectrometric detection of omega-oxidation products of aliphatic fatty acids in exhaled breath. *Anal Chem*. 2017;89(19):10329-10334. doi:10.1021/acs.analchem.7b02092

39. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012;30(10):918-920. doi:10.1038/nbt.2377

40. Hufsky F, Scheubert K, Böcker S. Computational mass spectrometry for small-molecule fragmentation. *Trends Anal Chem*. 2014;53:41-48. doi:10.1016/j.trac.2013.09.008

41. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrom*. 1994;5(9):859-866. doi:10.1016/1044-0305(94)87009-8

42. Gregorich M, Strohmaier S, Dunkler D, Heinze G. Regression with highly correlated predictors: Variable omission is not the solution. *Int J Environ Res Public Health*. 2021;18(8):4259 doi:10.3390/ijerph18084259

43. Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*. 5thed. Hoboken, New Jersey: John Wiley & Sons; 2012.

44. Mahajan U, Krishnan A, Malhotra V, Sharma D, Gore S. Predicting competitive weightlifting performance using regression and tree-based algorithms. In: Hassanien AE, Bhatnagar R, Darwish A, eds. *Advanced Machine Learning Technologies and Applications*. Singapore: Springer; 2020:397-415.

45. Gorshkov MV, Fornelli L, Tsybin YO. Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*. 2012;26(15):1711-1717. doi:10.1002/rcm.6289

46. Naito Y, Inoue M. Peak confluence phenomenon in Fourier transform ion cyclotron resonance mass spectrometry. *J Mass Spectrom Soc Jpn*. 1994;42(1):1-9. doi:10.5702/massspec.42.1

47. Kaufmann A, Walker S. Coalescence and self-bunching observed in commercial high-resolution mass spectrometry instrumentation. *Rapid Commun Mass Spectrom*. 2018;32(6):503-515. doi:10.1002/rcm.8054

48. Huang M-Z, Yuan C-H, Cheng S-C, Cho Y-T, Shiea J. Ambient ionization mass spectrometry. *Annu Rev Anal Chem*. 2010;3(1):43-65. doi:10.1146/annurev.anchem.111808.073702

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.