Journal of Breath Research

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Challenges in the identification and quantitation in on-line breath analysis

To cite this article: Timon Käser *et al* 2025 *J. Breath Res.* **19** 036002

Journal of Breath Research

PAPER

Challenges in the identification and quantitation in on-line breath analysis

Timon Käser, Stamatios Giannoukos and Renato Zenobi*

Laboratory of Organic Chemistry, ETH Zürich, Vladimir-Prelog-Weg 3, 8093 Zürich, Switzerland
* Author to whom any correspondence should be addressed.

E-mail: Zenobi@org.chem.ethz.ch

Keywords: online exhaled breath analysis, quantification, compound identification, tandem MS

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
10 January 2025

REVISED
25 March 2025

ACCEPTED FOR PUBLICATION
7 April 2025

PUBLISHED
23 April 2025

Abstract

The identification and quantitation of volatile organic compounds (VOCs) in exhaled human breath has attracted considerable interest due to its potential application in medical diagnostics, environmental exposure assessment, and forensic applications. Secondary electrospray ionization-mass spectrometry (SESI-MS) is a method capable of detecting thousands of VOCs. Nevertheless, most studies using SESI-MS for breath analysis have relied primarily on MS^1 measurements for identifications and quantification, which are susceptible to misassignments and errors. In this study, we targeted several endogenous compounds (C5 to C10 aldehydes, limonene and pyridine), known to occur in breath. These compounds were measured and quantified in exhaled breath from 12 volunteers over several days using three different acquisition methods: full scan, targeted selected ion monitoring and parallel reaction monitoring. These methods were used for identification and quantification by comparing with measurements of external standards. High-abundance features such as limonene and pyridine were successfully identified and quantified in exhaled human breath with all three methods, with MS^2 measurements supporting identification, albeit with limitations to separate between limonene and α-/β-pinene. For low-abundance features, the study highlights the challenges of false assignments in SESI-MS, even with MS^2 measurements. This was demonstrated in the case of aldehydes, which could not be reliably separated from isomeric ketones present in breath, leading to incorrect quantification.

1. Introduction

Medical diagnostics often require the precise quantification of biomarkers to distinguish between sick and healthy individuals. Precise quantification is also needed to define the therapeutic window for various drugs [1]. Traditionally, such quantification relies on samples like urine, blood, or even tissue biopsies, some of which can cause discomfort and pose risks to patients. In recent years, breath analysis has gained attention as a non-invasive alternative, avoiding these issues while still offering valuable diagnostic insights. Breath analysis involves the detection of inorganic gases, volatile organic compounds (VOCs), and aerosols, providing detailed information about physiological states and disease conditions. However, no VOC-based diagnostic test has entered clinical practice in recent decades, suggesting that more comprehensive studies are needed to reveal information contained in exhaled VOCs.

Clinical studies have explored the potential of breath analysis for detecting and monitoring specific substances, with some employing advanced techniques like secondary electrospray ionization-mass spectrometry (SESI-MS) [2]. SESI-MS offers enhanced sensitivity by efficiently ionizing molecules [3, 4], making it particularly suitable for the detection of trace compounds in breath. Despite its promise, SESI-MS faces challenges in reliably quantifying specific substances. The ionization efficiencies of SESI-MS are compound-specific and matrix-dependent, requiring calibration with gas standards, which is more labor-intensive than techniques like selected ion flow tube-mass spectrometry or proton

transfer reaction-mass spectrometry. These established methods use kinetic constants of analytes to calculate absolute concentrations directly from measured intensities [5].

Recent studies have investigated the quantification capabilities of SESI-MS, achieving detection limits down to parts per billion (ppb) or parts per trillion [3, 4, 6, 7]. However, to the best of our knowledge, no studies have yet conclusively demonstrated the successful quantification of substances in breath using SESI-MS. Current MS-based breath analyses rely primarily on MS^1 measurements, which, while effective for detecting a wide range of compounds, are susceptible to spectral overlap. For example, compounds like acetic acid and methyl formate [8], 1-propanol and 2-propanol [9], or different hydroxycarboxylic acids (2-hydroxy-, 3-hydroxy- and 8-hydroxyoctanoic acid) [10] are expected to exhibit identical spectral features. The number of overlapping features present remains unclear, often compromising the accurate identification and quantification of isomeric analytes.

In contrast, due to its high selectivity and sensitivity, multiple reaction monitoring is widely regarded as the gold standard for targeted analysis in liquid chromatography-mass spectrometry (LC-MS). High-resolution mass spectrometers, such as Orbitrap or Q-TOF instruments, enable even greater selectivity by monitoring multiple MS/MS fragmentations in a single scan, a method known as parallel reaction monitoring (PRM) [11]. While PRM offers significant advantages, it has not been widely applied in SESI-MS, probably because most studies have focused on exploratory, untargeted approaches. Therefore, little is known about the sensitivity and selectivity of MS^2 measurements in direct injection SESI-MS.

The research community emphasizes untargeted analysis using SESI-MS due to its capacity to detect hundreds or thousands of features in real time, making it a valuable tool for identifying potential biomarkers. Validated biomarkers can subsequently be measured in a targeted manner using sensors or other portable devices. However, we argue that integrating both targeted and untargeted analysis is crucial for enhancing the insights gained from SESI-MS and applying this knowledge to other techniques. This gap in understanding occurs because even when utilizing offline methods like GC-MS or LC-MS and observing the same mass as in online measurements, we cannot confirm the absence of other isomeric ions in online data. Therefore, there is a significant risk of misassigning artifacts as biomarkers by SESI-MS.

Given these gaps in the current research, our study aims to compare different measurement methods—specifically MS^1 and MS^2—in the context of breath analysis. We evaluate these methods in terms of their selectivity and sensitivity, which are critical for the accurate quantification of biomarkers in breath. Additionally, we discuss the potential and challenges of implementing targeted analysis approaches in breath analysis to advance the use of SESI-MS in medical diagnostics. By addressing these issues, this study contributes to developing more reliable and accurate non-invasive diagnostic tools based on breath analysis.

This work distinguishes between 'features' and 'compounds'. Features are peaks in the mass spectra that have not yet been identified. However, it is essential to note that even compounds may not always represent accurate identifications, as many compounds can only be 'identified' at a level 4 according to the classification by Schymanski *et al* [12]: level 4 identification means that the sum formula of the compounds was found, but no further information is available. For small molecules, even a level 2 identification based on MS^1 and MS^2 data but lacking supporting orthogonal information may be prone to errors. Despite these potential inaccuracies, we refer to these features as compounds but recommend to exercise caution in their interpretation.

In this study, we measured eight different compounds (limonene, pyridine, pentanal, hexanal, heptanal, octanal nonanal, and decanal) in breath. The aim was to attempt to identify and quantify these compounds using various MS^1 and MS^2 methods. These compounds were selected to evaluate the quantification capabilities of SESI for breath analysis. Limonene and pyridine show high intensity in breath, while aldehydes have lower intensities and therefore present greater challenges, although elevated concentrations of aldehydes have been found in patients with lung cancer and other diseases [13]. We evaluated the identification of these substances by comparing them to gas standard measurements. The same standards were used to prepare a calibration curve for each method and to quantify these compounds in breath. At the same time, quantitative comparisons between precursor ions and fragments were used to identify overlapping isomeric features in breath.

2. Experimental part

2.1. Reagents

Gas standards for the identification and quantification were prepared using analytical standards in the liquid phase. The following substances were utilized: pentanal (purity >97.5%), hexanal (purity $\geqslant$ 95%), heptanal (purity $\geqslant$ 97.0%), and (R)-(+)-limonene, purchased from Sigma Aldrich; nonanal (purity >95%) and decanal obtained from TCI; octanal (purity 99%) sourced from Merck; pyridine (purity 99.5%, extra dry) purchased from Acros and 3-pentanone (purity 99%) from abcr. Dilutions were made using Optima LC-MS grade water from Fisher Scientific. Optima LC-MS grade water (Fisher Chemicals) was also used to prepare a formic acid solution for the electrospray of the SESI source. For gas chromatography sample dilution, n-Hexane

(LiChrosolv LCMS grade) was purchased from Merck to dilute gas chromatography samples. The purity of the aldehydes used for quantification was confirmed via gas chromatography with a flame ionization detector (GC-FID), as detailed in the supporting information (S-1). The GC-FID results were used to calculate and apply purity adjustments during quantification.

2.2. Experimental setup

Online breath and standard gas measurements were conducted using a commercial SESI source (Fossil Ion Tech, Spain) coupled to an Orbitrap mass spectrometer (Q-Exactive plus; Thermo Fisher Scientific, USA). The sampling line of the SESI source was heated to 130 °C, while the SESI chamber was maintained at 90 °C. An electrospray solution of 0.1% formic acid in water was sprayed through a nano-electrospray capillary (inner diameter 20 μm) by applying a constant pressure of 0.8 bar to the vial containing the solution. The solution was ionized using a voltage of 3.5 kV, facilitating the ionization of gaseous compounds that enter the source. A constant pressure and flow into the source were maintained by a mass flow controller with an open connection to atmospheric pressure upstream of the sampling source. The Orbitrap mass spectrometer was operated with the following parameters: sheath gas flow of 15 a.u. (arbitrary unit), auxiliary gas of 2 a.u., S-lens RF amplitude of 50, and capillary temperature of 250 °C.

2.3. Gas standard production

Two systems were employed to produce gas standards in the ppb and low ppm ranges: reference compounds injected into evaporation chambers, or permeation tubes filled with pure reference compounds. Most of the gas standards were prepared using evaporation chambers, while the permeation tubes were only used for compounds where the gas concentrations generated by evaporation were too high relative to breath levels due to high volatility and high ionization efficiency.

Evaporation chambers generate gas standards by leveraging the vapor pressures of injected compounds. A small flow of nitrogen gas (1–10 ml min^{-1}) passes through the chamber, carrying the evaporated compounds and mixing with a larger flow of humidified nitrogen gas in mixing and dilution chambers. The humidified nitrogen flow was generated by bubbling nitrogen at 8 l min^{-1} through a wash bottle containing pure water (Optima, LC-MS grade, Fisher Chemicals) heated to 37 °C. The flow from the mixing chamber was then introduced into an adapter connected to the SESI source. The flow from the mixing chamber was directed into an adapter connected to the SESI source. The humidity level of the prepared standards was measured at approximately 33% relative humidity at 48 °C, which corresponds to 64% relative humidity at 37 °C. This range is similar to that found in exhaled breath, which is typically considered to be saturated; however, a recent study has indicated that it often falls below this level [14]. We also assume that the temperature difference between the breath and the standards can be disregarded after passing through the sampling line at 130 °C.

The process of generating gas standards from the evaporation chamber system has been described in detail in previous studies [3, 15, 16]. In this work, the evaporation chambers were maintained at 25 $\pm$ 1 °C. The system for mixing and dilution was constructed from passivated stainless-steel tubing and components to minimize adsorption of the compounds to the steel surfaces, as described in a prior study [17]. However, in this study no standard addition experiments were performed. For calibration, either 10 μl of pure compounds (e.g. limonene, as well as a mixture of aldehydes) or an aqueous solution (e.g. pyridine) were introduced into the chambers. For dilute aqueous solutions, the gas-phase concentration of the analyte was calculated using Henry's law, based on the Henry constant and the analyte concentration in the solution. Some compounds were not sufficiently soluble in water to produce gas-phase concentrations comparable to exhaled breath. Instead, these were introduced as pure or nearly pure substances, with gas-phase concentrations calculated using Raoult's law, because Henry's law is only valid for dilute solutions [18].

The physical principle used to calculate the gas concentration in the chambers is based on the evaporation of the substances and assumes that the gas phase is saturated, and the influence of diffusion is negligible. These assumptions align with those used in the transpiration method for measuring vapor pressures [19]. High nitrogen flow rates or a drop in the liquid-phase temperature could lead to undersaturation and reduced vapor pressure, potentially causing an overestimation of breath concentrations. However, these effects were not explored in this study as they do not impact the comparison of the measurement methods. The concentrations for all substances are given in table 1, and the calculations for determining the gas-phase concentrations are given in the supporting information (S-2).

For each measurement, a 10 mL/min flow was first passed through the gas standard generation system and diluted with the humidified nitrogen flow to record the system's background over a period of 120 s. A 10 ml min^{-1} flow was then passed through the evaporation chamber for 300 s to equilibrate and saturate the passivated lines. After a 120 s pause with no flow through the evaporation chamber, the calibration process began. Calibration was performed by

Table 1. Summary of all substances used in the evaporation chambers to prepare gas standards, including calculated partial pressures and gas-phase concentrations after dilution with humidified nitrogen. For pyridine (aqueous solution), Henry's law was applied to determine the partial pressure. For other substances, Raoult's law was used, based on the vapor pressure of the pure compound and the mole fraction in the solution.

substance	Molar mass (g mol^{-1})	Henry constant (mol m^3 Pa^{-1})	Concentration (mM)	$\chi_{solution}$	Vapor pressure (Pa)	P_i (Pa)
limonene	86.13	—	—	1	202	202
pyridine	79.10	1.1	0.5	—	—	0.45
pentanal	86.13	—	—	0.22	4844	1081
hexanal	100.16	—	—	0.19	1282	247.4
heptanal	114.19	—	—	0.17	383	64.4
octanal	128.21	—	—	0.15	151	22.9
nonanal	142.24	—	—	0.14	49	6.7
decanal	156.27	—	—	0.13	18	2.3

maintaining a 10 ml min^{-1} flow for 90 s, which was then reduced in a stepwise fashion to 8 ml min^{-1}, 6 ml min^{-1}, 4 ml min^{-1}, 2 ml min^{-1}, and 0 ml min^{-1} with each flow rate held for 90 s. The gas standard generation system was kept at 60 °C to prevent condensation. The humidified nitrogen flow was kept constant at 8 l min^{-1} throughout the experiments.

For quantifying 3-pentanone in breath, gas standards were prepared using permeation tubes. Approximately 500 mg of the substance were placed in an extruded PTFE tube sealed with PTFE stoppers at both ends. The tube, with a total length of 5 cm (without the plugs), was placed in a Calibration Generator (Owlstone, OVG-4), where it was heated at 40 °C in a stream of 50 ml min^{-1} of nitrogen. After overnight equilibration, the system was connected to the mixing and dilution setup used for the evaporation chambers. There, the 50 ml min^{-1} stream of gas standards was diluted with 8 l min^{-1} of humified nitrogen. Different gas concentrations could be produced from the gas generator by passing higher flows through the permeation oven and then splitting them. Gas concentrations were varied by adjusting the split flow (0, 100, 300, or 500 ml min^{-1}) of the OVG-4 and measure the exhaust flow with a float flowmeter.

The gas concentration was calculated from the permeation rate of 3-pentanone at 40 °C as described in the SI. The permeation rate was determined by measuring the weight loss of the permeation tube over several weeks. The measured permeation rate of 0.245 μg min^{-1} corresponds to a gas concentration of 1390 ppb through the permeation oven, which is similar to the permeation rate of 2-pentanone reported by Roberts *et al* [20].

2.4. MS methods

The rapid scanning speed of the Orbitrap mass spectrometer enables the sequential acquisition of different types of scans, allowing for direct comparison of various methods. In all methods, unless otherwise noted, three types of scans were employed. First, a full scan (FS) from 50 to 500 m/z with an injection time of 200 ms was performed, followed by a targeted selected ion monitoring (t-SIM) scan with a centered isolation window of 1 m/z. Lastly, a targeted PRM scan was conducted, also with a centered isolation window of 1 m/z. Figure S4 illustrates the timing and ion flow paths of these three methods.

The FS mode was measured with a resolution of 70 k and an automatic gain control (AGC) set to 3×10^6. The SIM mode was set to the same resolution of 70 k and an AGC of 5×10^4, and the PRM mode had a resolution of 35 k and an AGC of 2×10^5. The optimum collision energy for each substance was determined by optimizing the intensity of the dominant fragments across collision energies (CE) ranging from 10 to 100 eV. For quantification, a normalized CE (NCE) of 35 eV was selected for all aldehydes and limonene, while an absolute CE of 25 eV was chosen for pyridine. The use of an absolute CE for pyridine was necessary due to its higher chemical stability compared to the other substances, ensuring maximal fragmentation.

2.5. Identification and quantification of VOCs in exhaled human breath

Limonene, pyridine, pentanal, hexanal, heptanal, octanal, nonanal, and decanal were measured in exhaled human breath using all three methods (FS, SIM, and PRM) for identification and quantification. The quantification of these compounds was based on calibration curves recorded for all three methods under optimized conditions. Each breath measurement consisted of two sequences of five exhalations, allowing for the analysis of all eight compounds. To ensure sufficient data points were acquired during each exhalation, no more than four compounds were measured simultaneously. In total, 12 volunteers (between 27 and 41 years old) provided 25 breath samples over three days. For each participant, a new spirometry filter (Vyaire Medical, Germany) was used as a mouthpiece in front of the SESI source to prevent contamination from saliva or viruses. Volunteers were instructed to exhale at a flow rate of 8 l min^{-1}, monitored by a flow meter (EXHALION, Fossil Ion Tech, Spain) connected to the SESI source adapter.

Table 2. Summary of the ions and their m/z used to quantify the different compounds with the three different methods (FS, SIM and PRM). s.f.: sum formula.

		FS	SIM	PRM1		PRM2		PRM3	
Substance	Sum formula	$[M+H]^+$	$[M+H]^+$	s.f.	$[M]^+$	s.f.	$[M]^+$	s.f.	$[M]^+$
limonene	$C_{10}H_{16}$	137.1325	137.1325	C_8H_{13}	109.1012	C_7H_{11}	95.0855	C_6H_9	81.0698
pyridine	C_5H_5N	80.0495	80.0495	C_4H_5	53.0386	—	—	—	—
pentanal	$C_5H_{10}O$	87.0804	87.0804	C_5H_9	69.0699	C_4H_7	55.0542	—	—
hexanal	$C_6H_{12}O$	101.0961	101.0961	C_6H_{11}	83.0855	C_5H_9	69.0699	C_4H_7	55.0542
heptanal	$C_7H_{14}O$	115.1117	115.1117	C_7H_{13}	97.1012	C_5H_9	69.0699	C_4H_7	55.0542
octanal	$C_8H_{16}O$	129.1274	129.1274	C_6H_{11}	83.0855	C_5H_9	69.0699	C_4H_7	55.0542
nonanal	$C_9H_{18}O$	143.1430	143.1430	C_6H_{11}	83.0855	C_5H_9	69.0699	C_4H_7	55.0542
decanal	$C_{10}H_{20}O$	157.1587	157.1587	C_6H_{11}	83.0855	C_5H_9	69.0699	C_4H_7	55.0542
3-pentanone	$C_5H_{10}O$	87.0804	87.0804	C_5H_9	69.0699	C_4H_7	55.0542	C_4H_9	57.0699

Participants were instructed to abstain from eating or drinking anything other than water for 1 h before providing their breath samples. They were also asked to rinse their mouths with water immediately before the analysis and to refrain from using oral hygiene products or chewing gum during the hour leading up to the measurement.

A recently developed method called incremental quadrupole acquisition to resolve overlapping spectra (IQAROS) was applied to separate chimeric MS^2 spectra present in breath. This was done as described in the paper by Kaeslin *et al* for all substances in the breath of one volunteer [21].

2.6. Data processing and analysis

All spectra were acquired in reduced profile mode, a feature used by Thermo to minimize data storage by removing spectral noise. The data in this work were analyzed directly from the Thermo RAW files using the R package *rawrr*, which uses the vendor-provided API called RawFileReader to access all data stored in the RAW files [22].

Intensities were extracted as extracted ion chromatograms with a tolerance of 15 ppm. Unless otherwise specified, compounds were detected as their protonated molecular ion $[M+H]^+$. A list of all the ions used for calibration is provided in table 2. Scans corresponding to exhalations were identified using an algorithm that detects exhalations when the normalized TIC of the FS measurement exceeds 0.6. The TIC can be utilized because of its correlation with the measured exhaled flow. Whole exhalations were selected, as there were no significant differences in the average intensities of the chosen compounds compared to selecting only end-tidal exhalations.

The calibration curves were generated from the prepared gas standards as described in section 2.3. The intra-and inter-day variability of the quantification methods (FS, SIM, and PRM) was assessed by recording the calibration curves of all the compounds three times in one day and additionally on three separate study days. For each calibration step corresponding to different flows, the mean intensity of the last 45 s was calculated. The gas concentration for each step was calculated using either Henry's constant (for pyridine) or the vapor pressure (for limonene and the aldehydes) of the compounds, considering the dilution by the 8 l min^{-1} nitrogen flow. A weighted linear regression with a weight of $1/x^2$ was used for calibration, where x stands for the gas concentration. For certain compounds, such as limonene and some aldehydes, extrapolation was used to calculate breath concentrations. This approach avoided the need for additional dilution steps of the gas stream or substance dilution in organic solvents, which would have required vapor pressure measurements in those solvents and expanded the scope of this study.

For identification, the averaged MS^2 spectra from the scans of one exhalation were used. These spectra were either extracted as .mgf files for the IQAROS measurements or as .csv files for the database search using SIRIUS [23]. The database search workflow with SIRIUS included: calculation of the molecular formula of the precursor and a search for it in the human database, a fingerprint prediction using CSI:FingerID [24], and a CANOPUS [25] search to predict the compound class with the help of ClassyFire [26] and NPClassifier [27].

3. Results and discussion

3.1. Identification of compounds with SIM and PRM spectra

All selected compounds (limonene, pyridine, and the selected aldehydes) were detected as their protonated molecular ion $[M+H]^+$ in the FS and SIM modes in the breath measurements of the volunteers. The high mass accuracy of the Orbitrap enabled confirmation of the sum formula of the compounds detected in breath. The PRM mode provided additional structural information, but this is dependent on the purity of the MS^2 spectra. Chimeric MS^2 spectra, where the target precursor contributes less than 50% of all the ions used for fragmentation, usually cannot be reliably identified through database comparisons. The IQAROS method, on the other hand, can only be applied to measurements where the precursor is greater than at least 5% of the total ion current (TIC)

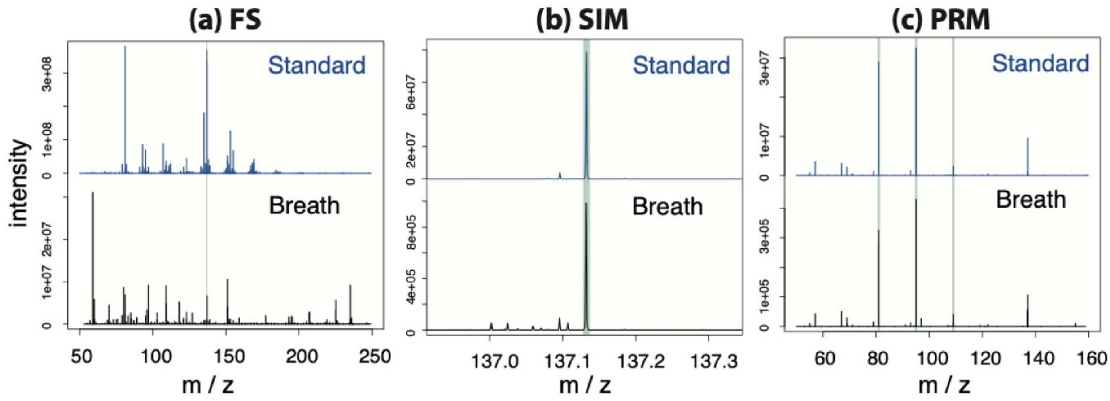

Figure 1. Mass spectra of limonene in gas standard and breath recorded in three different scan modes: (a) full scan (FS), (b) targeted selected ion monitoring (t-SIM) and (c) parallel reaction monitoring (PRM) modes. The protonated molecular ion of limonene (m/z 137.1325) is highlighted in (a) and (b) and the fragments selected for PRM (m/z 81.0697, 95.0851 and 109.1012) are highlighted in (c).

in the selected mass range and is still visible in the MS^2 spectra [21].

For limonene and pyridine, the features detected in breath accounted for more than 50% of the ions within the SIM window, allowing these spectra to be used directly for database searches via SIRIUS (version 5.8.6, see SI S-7 for more information). Pyridine was clearly identified with 90% agreement in the database. For limonene, a definitive identification was not achieved, but the CANOPUS search classified it as a monoterpene with a probability of 65%.

For pentanal, an intensity ratio of more than 50% was observed in the breath of some volunteers, enabling further identification. The CANOPUS search confirmed it as a carbonyl compound with 97% probability. Proposed structures included the ketones 2-pentanone and 3-pentanone and the aldehydes pentanal, isopentanal, or 2-formylbutane. While CSI:FingerID scores favored the ketones, the similarity scores were better for the aldehydes. All other aldehydes could not be identified from their chimeric MS^2 spectra. Although IQAROS was tested to separate the chimeric spectra of the aldehydes, the aldehydes were not visible as precursors in the MS^2 spectra, rendering IQAROS inapplicable.

All compounds were quantified to compare their concentrations with reported literature values and to examine the ratios between precursor ions and fragments. This comparison was achieved by analyzing breath samples alongside gas standard measurements of the proposed compounds. Fragment masses used for quantification were derived from measurements of pure compounds. For example, figure 1 illustrates the measurement of limonene in breath and gas standards using all three methods (FS, SIM, and PRM). The masses used for the quantification are highlighted. In the FS measurement, we observe the same mass (m/z = 81.0697) as in the fragmentation spectra, indicating that in-source fragmentation may occur. This is discussed further in section 3.4. The mass spectra of the other compounds are provided in the SI (S-6).

3.2. Calibration curves for all three methods

Quantification of the compounds by all three methods enabled investigation of the precursor-to-fragment ion ratios in breath and comparison with gas standard measurements. For this purpose, a calibration curve was constructed for each standard and method. The averaged intensity of the molecular ions or fragments was plotted against the gas standard concentrations, and a separate weighted linear regression was performed for each method. An example of the calibration curve for limonene is shown in figure 2, with additional calibration curves provided in the supporting information (SI-9). The origin (zero concentration of reference gas) was excluded in the regression analysis because it is difficult to measure an accurate background with this system due to the long washout times. Figures 2 and S16–S23 demonstrate a linear dependence of the signal over the concentration range used in this study (to prepare higher dynamic range see SI). While slight deviations from linearity were observed for some compounds, these deviations were consistent across all three methods. Consequently, the deviations do not impact the relative comparison of the FS, SIM, and PRM methods.

The intra- and inter-day variances of the calibrations are included in the supporting information (table S4), along with the calculated limits of detection and limits of quantification (table S5). The inter-day variance, which reaches up to 30%, is larger than the intra-day variation. This indicates that there were some challenges to generate standards, likely due to the need for quite long equilibration times or potential batch effects of the electrospray. In general, we used the calibration on the same day as measuring volunteers' breath. For limonene, which required the

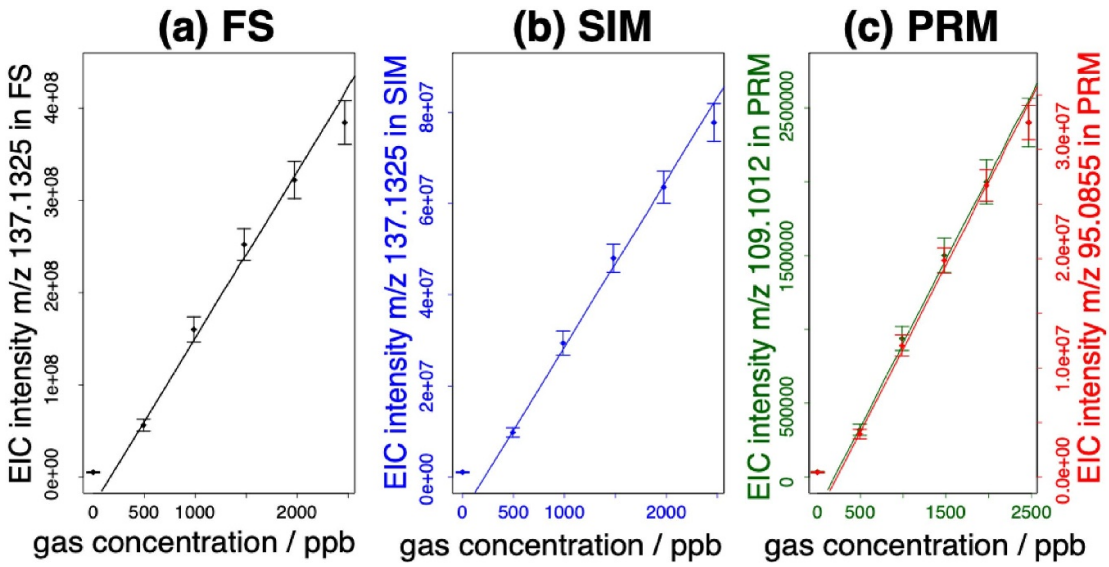

Figure 2. Calibration curves for limonene recorded in three different scan modes: (a) full scan (FS), (b) targeted selected ion monitoring (t-SIM) and (c) parallel reaction monitoring (PRM) modes. Limonene was introduced as gas standards at different concentrations. For each concentration, the mean intensity of the extracted ion chromatogram (EIC) of m/z 137.1325 was calculated for the FS and SIM scan mode. For the PRM mode, the mean intensity of the EIC of the two fragments m/z 109.1012 and 95.0855 were taken. The mean of four different measurement on the same measurement day are shown and used for the weighted linear regression (weights = $1/x^2$). The origin was not included in the linear regression. The error bars show the mean standard error calculated by error propagation.

longest equilibration time, this was not possible, and we used the calibration from a different day.

Daily variations of the electrospray introduce some errors in the calculation of the concentrations for limonene due to the fact that the calibration was measured on a different day. However, this does not affect the comparison of the three different methods (FS, SIM, and PRM). For this, it is only necessary that all three methods are measured at the same concentration rather than requiring exact absolute values. Consequently, the daily variations in the electrospray result in a systematic error that is consistent across all methods.

3.3. Quantification of limonene and pyridine in human exhaled breath

Quantification in exhaled breath was performed using the same protonated molecular ions and fragments as those used in the calibrations. The TIC of the FS measurement was aligned with the flow recorded by the Exhalion to detect the exhalation. The intensities of the characteristic ions were averaged across all scans for each scan mode. One measurement cycle, including all three scan modes, required around 500 ms. Previous experiments have demonstrated a variation in TIC between scans within 5% for breath measurements. Averaging multiple scans reduces this variation, allowing us to assume that the concentration remains constant during one measurement cycle in a consistent exhalation.

The concentrations for limonene and pyridine in breath, calculated from the measurements, are shown in figure 3. Limonene was detected in breath at a concentration of around 250 ppb and pyridine at a concentration of around 3 ppb. Since most limonene measurements were below the lowest calibration point, its concentrations were estimated by extrapolation. We observed no significant differences in the concentration ranges for either substance measured using FS, SIM, or PRM, suggesting that all three methods are suitable for quantifying these substances in breath. The consistent ratio of concentrations observed by SIM and PRM suggests that there are no isomeric substances with different fragments or isobaric substances with the same fragments present in breath. The presence of either of these would lead to a different fragment-to-precursor ratio. Only an isomeric interference that creates a very similar fragment pattern could maintain a constant ratio and hence go undetected. However, this limitation cannot be addressed in direct injection mass spectrometry without employing an orthogonal separation technique.

The concentration of limonene is likely overestimated compared to the literature, where it was reported to be about 1.5 ppb [28, 29]. This overestimation is probably due mostly to the presence of pinene in breath (about 1.2 ppb [28]; pinene has the same molecular weight and fragmentation pattern as limonene (shown in S-10) and a lower ionization efficiency than limonene (see in SI for α-pinene). The compounds can therefore not be distinguished by direct-injection mass spectrometry [30]. Consequently, we used limonene as a proxy for the total concentration of these monoterpenes in breath. Additional contributions to the deviation from the

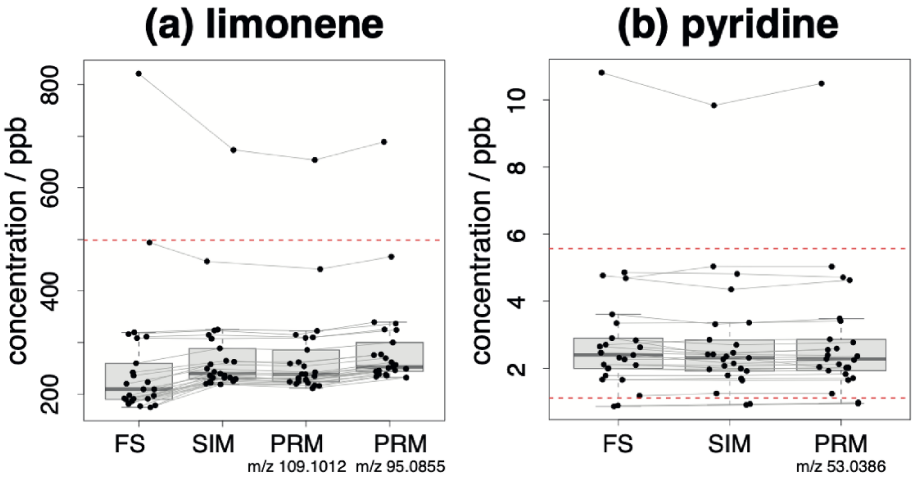

Figure 3. Box plots of the concentrations of (a) limonene (used as a proxy for monoterpenes) and (b) pyridine measured in breath of volunteers with all three methods. Points outside the box plot are outliers. The dashed red lines indicate the lowest and highest calibration points used in the study. The grey lines indicate concentrations calculated from the same breath measurement. Details to the calculation of the box plots and the outliers are given in the SI. Only two of the three fragments are shown for limonene.

expected concentration of limonene include the calibration not being performed on the same day, errors associated with extrapolation, and possible undersaturation of the gas flow through the evaporation chambers. Further research is necessary to validate the assumption of saturation used in calculating the gas concentrations of the standards.

A direct comparison between the individual concentration measured with SIM and PRM is shown in the Bland–Altman plots in figure 4. These plots illustrate the relationship between the differences in concentration measurements from the two methods and their average values for both limonene and pyridine. For limonene, we observe that the concentrations measured with PRM are slightly higher on average, presumably because breath is a mixture of limonene and α/β-pinene, and there are minor differences in the ratio of fragments for both substances. This trend appears more pronounced at higher concentrations (>400 ppb), possibly reflecting the recent consumption of foods by participants that contain one of these compounds. However, more datapoints would be needed to confirm this trend. For pyridine, both concentrations are distributed around the same mean. Further Bland–Altman plots for the other compounds tested are shown in the SI.

3.4. Potential in-source fragmentation and its influence on the quantification and identification

During the measurement of limonene, we observed large intensities of the fragment $C_6H_9^+$ in the FS measurement. Although other fragments were detected, our focus is on this fragment due to its predominant intensity. The fragment is probably generated due to in-source fragmentation, although SESI is thought to be a soft ionization method. Recent work, however, has shown that these SESI conditions may not be as soft as previously assumed [6, 31–33]. To evaluate its influence on the quantification, we compared the ratio of the fragment to the protonated molecular ion. As shown in figure 5, the ratios are similar between the standard and breath measurements, indicating that the calculated concentrations are accurate and not significantly influenced by in-source fragmentation. In a few breath measurements, the ratio was higher, implying a possible overlap with another breath feature.

The FS measurement of aldehydes in the standard mixture also reveals high intensities of potential fragments, with the dominant fragment for shorter aldehydes resulting from the loss of a water molecule. The features $C_nH_{2n-1}O_2^+$ are also found in the breath and standard measurements, but it remains to be fully understood whether these arise from in-source reactions or potential oxidation products of the aldehydes.

Further work is needed to investigate the influence of in-source fragmentation on other features in breath. The in-source fragments may be misidentified as other compounds, leading to erroneous conclusions, or they may overlap with other features and thus falsify the intensity of these compounds. Nevertheless, we were able to show that the quantification of monoterpenes is not influenced by the presence of in-source fragmentation.

3.5. Quantification of carbonyl compounds in exhaled human breath

In principle, one should be able to determine the breath concentrations of the six different aldehydes similarly to limonene and pyridine. Table S6 summarizes the measured concentrations for all the aldehydes. All calculations were performed using the data from the SIM measurements. It was observed that the concentration calculated from FS data is smaller compared to that from SIM data for all aldehydes, except

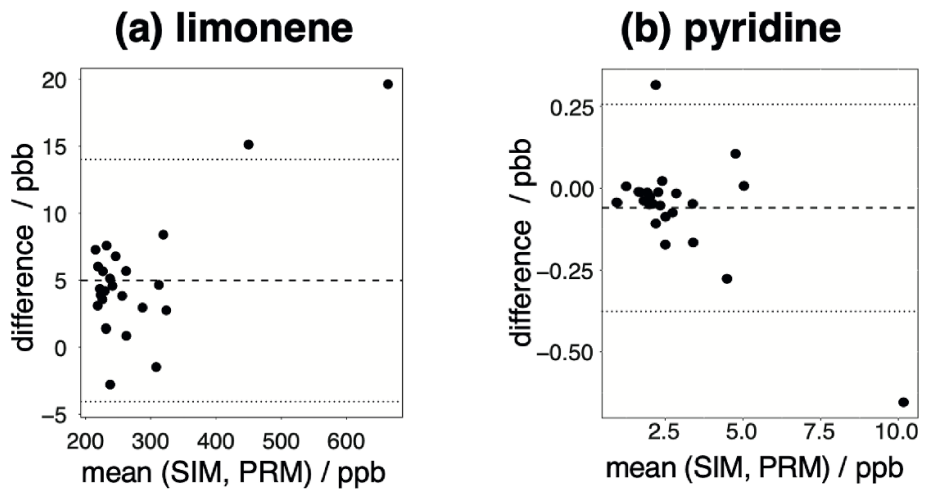

Figure 4. Bland–Altman plots to compare the agreement of the breath concentration calculated with SIM and PRM of (a) limonene and (b) pyridine. The dashed line shows the mean of all measurements, illustrating if we have a bias between the two methods. On average, we measure slightly higher concentration with SIM than PRM for limonene. The dotted lines represent the agreement limits (± 1.96 * standard deviation) in which the difference of 95% of all measurements is.

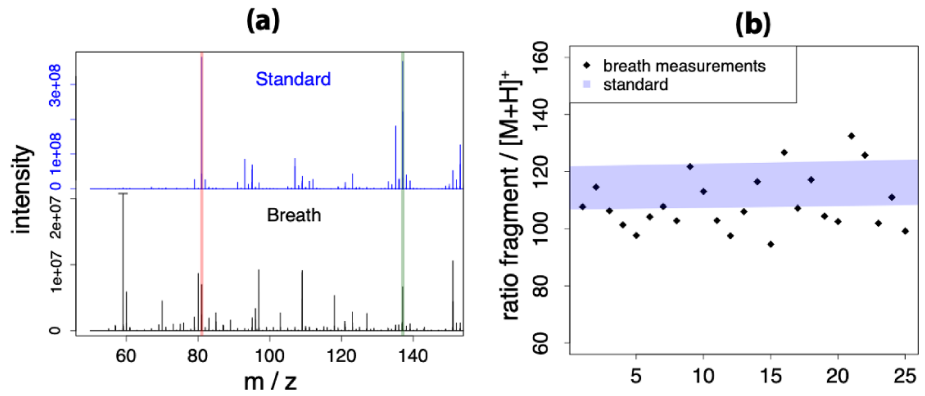

Figure 5. (a) Averaged mass spectrum of breath measurement from one volunteer compared to a gas standard measurement of limonene. Both mass spectra show a full-scan measurement. The signal at m/z = 137.1325 corresponding to protonated limonene, $[M + H]^+/(C_{10}H_{17}^+)$, is highlighted in green. The signal at m/z = 81.0699 corresponding to the most intense fragment of limonene ($C_6H_9^+$) is highlighted in red. (b) Ratio of the intensity of fragment ($C_6H_9^+$) to the intensity of $[M + H]^+$ shown for all breath measurements. The ratio of the standard measurement is displayed as the blue background.

pentanal. This is probably due to ion competition inside the C-Trap of the Orbitrap instrument, which is more pronounced for less abundant compounds.

However, of greater concern was that the measured concentrations were much higher than those reported in the literature. For example, we measured about 4.3 ppm for pentanal, compared to literature measurements of less than 10 ppb, as reported in a study with 700 participants measured by LC-MS [34]. Similar overestimation of the breath concentration was found for all 6 aldehydes measured, and if some of the fragments were used for quantitation, the discrepancy was sometimes even larger. We propose that this is due to an overlap of signals of the tested aldehydes with signals from isomeric substances. For example, Xie *et al* reported the presence of saturated ketones (pentanone $C_5H_{10}O$, hexanone $C_6H_{12}O$, heptanone $C_7H_{14}O$ and octanone $C_8H_{16}O$) in breath, whose molecular ion signals overlap with those of the aldehydes (pentanal $C_5H_{10}O$, hexanal $C_6H_{12}O$, heptanal $C_7H_{14}O$ and octanal $C_8H_{16}O$) [35]. In fact, about 10 times higher concentrations were reported for these ketones than for the aldehydes (e.g. 18.9 ppb for pentanone vs. 2.6 ppb for pentanal).

To validate whether most of the signals are due to ketones, we prepared gas standards of 3-pentanone with a permeation oven to perform a calibration (shown in SI). If we use this calibration to quantify the observed signal intensity of $C_5H_{10}O$, measured as $[M + H]^+$ a mean breath concentration of 9.4 ppb pentanone was calculated, which is within the concentration range (1.8–31.3 ppb) for pentanone in the breath of 20 volunteers [35]. Nevertheless, the quantification of ketones will also never be correct due to the presence of aldehydes and maybe other isomeric compounds. Therefore, it is recommended to

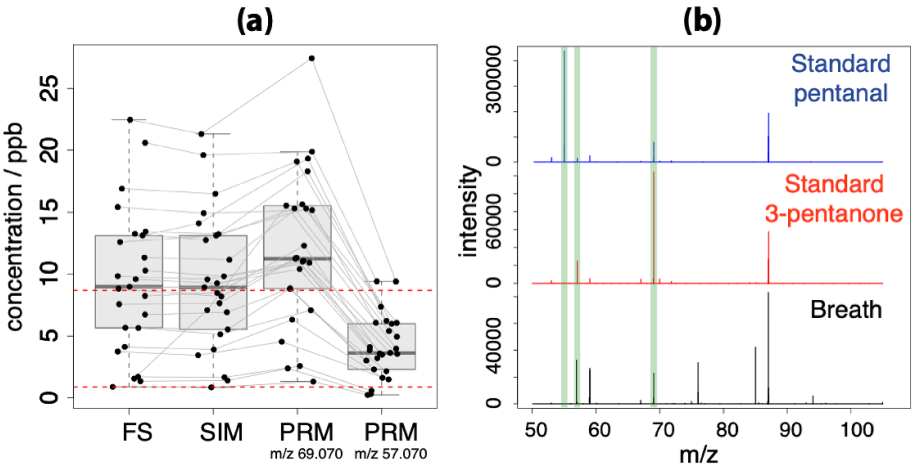

Figure 6. (a) Box plots of the concentrations of 3-pentanone measured in the breath of volunteers with all three methods. 3-pentanone was used as a quantitative marker to estimate the concentration of carbonyl compounds in breath. The dashed red lines indicate the lowest and highest calibration points used in the study. Two different fragments of 3-pentanone were chosen from the standard measurement and the concentration was calculated for each of the fragment. The grey lines connect concentrations calculated from the same breath measurement. (b) Averaged mass spectra of PRM measurement from the precursors in the m/z 87.1 $\pm$ 0.5 range. The spectrum from breath of one volunteer is shown in in black. The averaged spectra of two compounds at the highest concentration of the calibration curve are shown in blue and red. The three major fragments m/z 69.0699, m/z 55.0542 and m/z 57.0699 are highlighted in green.

use the corresponding signals to determine a summary parameter for carbonyls, rather than for individual compounds.

The comparison of concentrations calculated for 3-pentanone with the different measurement methods is shown in figure 6(a). This data also clearly confirms the presence of isomers because different concentrations are obtained depending on the fragment chosen. For instance, concentrations derived from the fragment with m/z = 69.070 appear slightly overestimated, while those based on the fragment with m/z = 57.070 are underestimated.

PRM is known to improve the selectivity of quantification in LC-MS experiments. Therefore, we also tested whether it could be used to discriminate between ketones and aldehydes in SESI-MS. However, the measurements of pentanal and 3-pentanone have similar MS^2 spectra (shown in figure 6(b)) and share some fragments, which is also the case for the other aldehydes measured in this study. There are some differences that could allow the separation of the aldehydes and ketones (e.g. $C_4H_9^+$ fragment with m/z 57.0699 for pentanone). Therefore, it could be that a calibration based on this fragment would afford the true concentration of pentanone. However, to prove this we would have to validate that that there is no other isobaric substance present that generates this fragment. This is mostly not possible: as an example, we observe features with the sum formula $C_nH_{2n-1}O_2^+$ in the gas standards of the aldehydes as well as in the breath measurements. They seem to have a similar fragmentation as the corresponding ketones.

Further research is needed to investigate all the fragmentation pathways of these carbonylic compounds and to attempt to separate them by SESI-MS. Nevertheless, our findings highlight the importance of careful fragment selection for quantification to ensure that all potential isobaric features in the breath matrix are taken into account. Only by confirming that target compounds do not share overlapping fragments, we can achieve more accurate quantification.

4. Conclusions

To date, relatively few studies have employed SESI to quantify substances in human breath or used on-line MS^2 spectra for identification. In this work, we show how to enhance the identification of monoterpenes and pyridine in human breath and successfully quantify them in the breath of 12 volunteers. Since on-line mass spectrometry cannot distinguish between limonene and α-/β-pinene, the overall concentration of these monoterpenes was estimated using limonene as a proxy. No significant differences were observed for any of these compounds when measured by FS, selected ion monitoring or PRM. The use of SIM in combination with PRM provided improved accuracy of compound identification and demonstrated the absence of interfering isomeric substances present in breath that have a different fragmentation. In addition, we showed that these compounds are not affected by ion competition and can be quantified in the FS regularly used in SESI-MS.

During the measurement of limonene, in-source fragmentation was observed, as reported in previous work [32, 33]. Similarly, in-source fragmentation was detected in breath samples, suggesting that its extent is equivalent when comparing standard samples and

exhaled breath samples. Consequently, the reported concentrations of limonene remain unaffected by in-source fragmentation, ensuring the reliability of our quantification approach.

In addition to monoterpenes and pyridine, aldehydes including pentanal, hexanal, heptanal, octanal, nonanal, and decanal were attempted to be identified and quantified. Their protonated molecular ions $[M + H]^+$ were detected. However, their low abundance and the overlap with other isobaric compounds made it impossible to prove their presence and confirm the identification by comparing with the fragmentation pattern found in standard measurements. The concentrations measured for these aldehydes by FS and SIM were significantly higher than those reported in the literature, probably due to overlap with isomeric ketones. Similar fragmentations of aldehydes and ketones, together with other overlapping features, prevent the separation of the compounds by direct-injection mass spectrometry. The MS^2 measurements underscore the importance of selecting specific features without interference from other compounds present in the same MS^1 window. This highlights the problems associated with the presence of isomeric compounds in breath, and the challenge of using MS^2 measurements to quantify compounds in exhaled breath, because compounds with similar fragmentation patterns may coexist within the same isolation window. Addressing this challenge will require further research to systematically investigate compounds in exhaled breath that share fragment ions, a task that is further complicated by the presence of in-source fragments.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.3929/ethz-b-000714426.

Acknowledgments

The research for this project was funded by Agroscope (NutriExhalomics project, Grant Agreement No. 15668) and by the Heidi Ras Foundation.

Ethics

The study was approved by the Ethics Committee of the ETH Zürich (EK-2024-N-187) and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained for all volunteers.

ORCID iDs

Timon Käser https://orcid.org/0009-0005-4056-9159
Stamatios Giannoukos https://orcid.org/0000-0002-5066-533X
Renato Zenobi https://orcid.org/0000-0001-5211-4358

References

[1] Schmidt F, Osswald M, Zenobi R and Kohler M 2022 Monitoring drug pharmacokinetics *Volatile Biomarkers for Human Health: From Nature to Artificial Senses* (Royal Society of Chemistry) ch 17, pp 323–46
[2] Hu B 2023 Mass spectrometric analysis of exhaled breath: recent advances and future perspectives *TRAC Trends Anal. Chem.* **168** 117320
[3] Wüthrich C, Fan Z, Vergères G, Wahl F, Zenobi R and Giannoukos S 2022 Analysis of volatile short-chain fatty acids in the gas phase using secondary electrospray ionization coupled to mass spectrometry *Anal. Methods* **15** 553–61
[4] Liu C, Zeng J, Sinues P, Fang M, Zhou Z and Li X 2021 Quantification of volatile organic compounds by secondary electrospray ionization-high resolution mass spectrometry *Anal. Chim. Acta* **1180** 338876
[5] Bruderer T, Gaisl T, Gaugg M T, Nowak N, Streckenbach B, Müller S, Moeller A, Kohler M and Zenobi R 2019 On-line analysis of exhaled breath *Chem. Rev.* **119** 10803–28
[6] Wolf J-C, Schaer M, Siegenthaler P and Zenobi R 2015 Direct quantification of chemical warfare agents and related compounds at low ppt levels: comparing active capillary dielectric barrier discharge plasma ionization and secondary electrospray ionization mass spectrometry *Anal. Chem.* **87** 723–9
[7] Martínez-Lozano P, Rus J, Fernández de la Mora G, Hernández M and Fernández de la Mora J 2009 Secondary electrospray ionization (SESI) of ambient vapors for explosive detection at concentrations below parts per trillion *J. Am. Soc. Mass Spectrom.* **20** 287–94
[8] Pysanenko A, Španěl P and Smith D 2009 Analysis of the isobaric compounds propanol, acetic acid and methyl formate in humid air and breath by selected ion flow tube mass spectrometry, SIFT-MS *Int. J. Mass Spectrom.* **285** 42–48
[9] Czippelová B, Nováková S, Šarlinová M, Baranovičová E, Urbanová A, Turianiková Z, Krohová J Č, Halašová E and Škovierová H 2024 Impact of breath sample collection method and length of storage of breath samples in Tedlar bags on the level of selected volatiles assessed using gas chromatography-ion mobility spectrometry (GC-IMS) *J. Breath Res.* **18** 036004
[10] Gaugg M T 2019 On-line breath metabolomics with ambient high-resolution mass spectrometry *ETH Zürich* (https://doi.org/https://doi.org/10.3929/ethz-b-000351718)
[11] Gallien S, Duriez E, Crone C, Kellmann M, Moehring T and Domon B 2012 Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer *Mol. Cell. Proteomics* **11** 1709–23
[12] Schymanski E L, Jeon J, Gulde R, Fenner K, Ruff M, Singer H P and Hollender J 2014 Identifying small molecules via high resolution mass spectrometry: communicating confidence *Environ. Sci. Technol.* **48** 2097–8
[13] Floss M A, Fink T, Maurer F, Volk T, Kreuer S and Müller-Wirtz L M 2022 Exhaled aldehydes as biomarkers for lung diseases: a narrative review *Molecules* **27** 5258
[14] Mansour E, Vishinkin R, Rihet S, Saliba W, Fish F, Sarfati P and Haick H 2020 Measurement of temperature and relative humidity in exhaled breath *Sens. Actuators* B **304** 127371
[15] Statheropoulos M, Pallis G C, Mikedi K, Giannoukos S, Agapiou A, Pappa A, Cole A, Vautz W and Thomas C L P 2014 Dynamic vapor generator that simulates transient odor

emissions of victims entrapped in the voids of collapsed buildings *Anal. Chem.* **86** 3887–94
[16] Giannoukos S, Marshall A, Taylor S and Smith J 2017 Molecular communication over gas stream channels using portable mass spectrometry *J. Am. Soc. Mass Spectrom.* **28** 2371–83
[17] Wüthrich C, Käser T, Zenobi R and Giannoukos S 2024 Internal standard addition system for online breath analysis *Anal. Chem.* **96** 10871–6
[18] Sander R 2023 Compilation of Henry's law constants (version 5.0.0) for water as solvent *Atmos. Chem. Phys.* **23** 10901–2440
[19] Kvande H and Wahlbeck P G 1976 Theory for the Determination of Vapour Pressures by the Transpiration Method *Acta Chem. Scand.* A **30** 297–302 (available at: https://www.academia.edu/112051038/Theory_for_the_Determination_of_Vapour_Pressures_by_the_Transpiration_Method)
[20] Roberts I J, Carpenter L J, Shaw M D and Langford V S 2022 Selected Ion Flow Tube—Mass Spectrometry (SIFT-MS) study of the reactions of H3O+, NO+ and O2+ with a range of oxygenated volatile organic carbons (OVOCs) *Int. J. Mass Spectrom.* **479** 116892
[21] Kaeslin J and Zenobi R 2022 Resolving isobaric interferences in direct infusion tandem mass spectrometry *Rapid Commun. Mass Spectrom.* **36** 1–11
[22] Kockmann T and Panse C 2021 The rawrr R package: direct access to orbitrap data and beyond *J. Proteome Res.* **20** 2028–34
[23] Dührkop K, Fleischauer M, Ludwig M, Aksenov A A, Melnik A V, Meusel M, Dorrestein P C, Rousu J and Böcker S 2019 SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information *Nat. Methods* **16** 299–302
[24] Dührkop K, Shen H, Meusel M, Rousu J and Böcker S 2015 Searching molecular structure databases with tandem mass spectra using CSI:FingerID *Proc. Natl Acad. Sci. USA* **112** 12580–5
[25] Dührkop K *et al* 2020 Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra *Nat. Biotechnol.* **39** 462–71
[26] Djoumbou Feunang Y *et al* 2016 ClassyFire: automated chemical classification with a comprehensive, computable taxonomy *J. Cheminform.* **8** 1–20
[27] Kim H W, Wang M, Leber C A, Nothias L-F, Reher R, Kang K B, van der Hooft J J J, Dorrestein P C, Gerwick W H and Cottrell G W 2021 NPClassifier: a deep neural network-based structural classification tool for natural products *J. Nat. Prod.* **84** 2795–807
[28] Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M and Amann A 2013 Blood and breath levels of selected volatile organic compounds in healthy volunteers *Analyst* **138** 2134–45
[29] O'Hara M E, Fernández Del Río R, Holt A, Pemberton P, Shah T, Whitehouse T and Mayhew C A 2016 Limonene in exhaled breath is elevated in hepatic encephalopathy *J. Breath Res.* **10** 046010
[30] Uller M M *et al* 2009 Atmospheric measurement techniques MS/MS studies for the selective detection of isomeric biogenic VOCs using a townsend discharge triple quadrupole tandem MS and a PTR-Linear Ion Trap MS (available at: www.atmos-meas-tech.net/2/703/2009/)
[31] Kaeslin J, Wüthrich C, Giannoukos S and Zenobi R 2022 How soft is secondary electrospray ionization? *J. Am. Soc. Mass Spectrom.* **33** 1967–74
[32] Gisler A, Lan J, Singh K D, Usemann J, Frey U, Zenobi R and Sinues P 2020 Real-time breath analysis of exhaled compounds upon peppermint oil ingestion by secondary electrospray ionization-high resolution mass spectrometry: technical aspects *J. Breath Res.* **14** 046001
[33] Zeng J, Christen A, Singh K D, Frey U and Sinues P 2022 Comparison of plasma ionization- and secondary electrospray ionization- high-resolution mass spectrometry for real-time breath analysis *Chimia* **76** 127–32
[34] McCartney M M *et al* 2020 Breath carbonyl levels in a human population of seven hundred participants *J. Breath Res.* **14** 046005
[35] Xie Z, Morris J D, Mattingly S J, Sutaria S R, Huang J, Nantz M H and Fu X-A 2023 Analysis of a broad range of carbonyl metabolites in exhaled breath by UHPLC-MS *Anal. Chem.* **95** 4344–52