Contents lists available at ScienceDirect

Atmospheric Environment

A new processing scheme for ultra-high resolution direct infusion mass spectrometry data

Arthur T. Zielinski[a], Ivan Kourtchev[a], Claudio Bortolini[b], Stephen J. Fuller[a], Chiara Giorio[a], Olalekan A.M. Popoola[a], Sara Bogialli[b], Andrea Tapparo[b], Roderic L. Jones[a], Markus Kalberer[a,*]

[a] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[b] Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, 35131 Padova, Italy

GRAPHICAL ABSTRACT
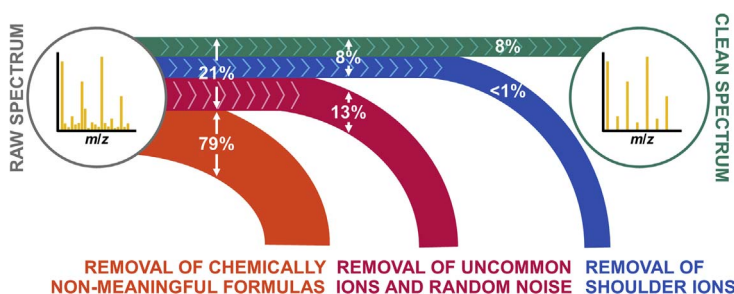
ARTICLE INFO

ABSTRACT

High resolution, high accuracy mass spectrometry is widely used to characterise environmental or biological samples with highly complex composition enabling the identification of chemical composition of often unknown compounds. Despite instrumental advancements, the accurate molecular assignment of compounds acquired in high resolution mass spectra remains time consuming and requires automated algorithms, especially for samples covering a wide mass range and large numbers of compounds. A new processing scheme is introduced implementing filtering methods based on element assignment, instrumental error, and blank subtraction. Optional post-processing incorporates common ion selection across replicate measurements and shoulder ion removal. The scheme allows both positive and negative direct infusion electrospray ionisation (ESI) and atmospheric pressure photoionisation (APPI) acquisition with the same programs. An example application to atmospheric organic aerosol samples using an Orbitrap mass spectrometer is reported for both ionisation techniques resulting in final spectra with 0.8% and 8.4% of the peaks retained from the raw spectra for APPI positive and ESI negative acquisition, respectively.

1. Introduction

Obtaining correct elemental composition of unknown molecules by mass spectrometry is a challenge despite advances in instrumentation and data processing algorithms (Kind and Fiehn, 2007). Ultra-high-resolution mass spectrometry (UHRMS), coupled with soft ionisation techniques, most commonly electrospray ionisation (ESI), can provide a detailed molecular composition for a large, complex sample (Nizkorodov et al., 2011; Kourtchev et al., 2014a) being able to identify many distinct peaks at a given nominal mass.

Manual data processing and formula assignment is extremely time consuming (Kind and Fiehn, 2007; Koch et al., 2007) so automatic

algorithms have been developed that generally include noise elimination and blank subtraction steps (Wong et al., 2009; Danger et al., 2013). Noise filtering and blank subtraction is challenging for analysis in direct infusion without prior chromatographic separation as the ion intensities may not be directly related to the concentration of the molecules in the sample (Kourtchev et al., 2014b).

Constraints on allowed chemical elements and number of atoms are used when assigning molecular formulae due to chemical reasons and computational limits. In a molecule containing only carbon and hydrogen the "rule of 13" can be used to limit the number of carbon atoms, in which the nominal mass is divided by 13 and the numerator gives the number of carbon atoms and the remainder gives the number of hydrogens (Bright and Chen, 1983; Koch et al., 2005). However, natural organic matter is mainly composed of C, H, O and N with minor contributions from S and P, the latter being a quantitatively non-significant component and often not considered (Koch et al., 2007; Wozniak et al., 2008; Ohno and Ohno, 2013). The number of possible solutions for an elemental formula increases largely if non-oxygen heteroatoms are taken into consideration for calculation. Calculating unique elemental compositions is not always possible (Kind and Fiehn, 2007) when acquiring data with high mass accuracy and resolution, especially as increasing the molecular mass of analytes increases the number of possible molecular formula assignments exponentially (Koch et al., 2007). In addition, because mass spectrometry does not directly provide structural information, these molecular formulae may represent any of several structural isomers.

In order to automatically constrain the large number of possible candidates, rules have been developed to select the most likely and chemically meaningful molecular formulae (Kind and Fiehn, 2007). An important constraint for restricting formulae to those that are likely to exist in nature is including element ratios, especially the H/C ratio which, in most cases does not exceed $H/C > 3$ (Kind and Fiehn, 2007). Similar restrictions can be put on the O/C (taking acidic polysaccharides as the upper limit for molecular oxygen content $O/C < 1.3$) (Koch et al., 2007) and other heteroatoms to carbon ratios (Kourtchev et al., 2013; Ohno and Ohno, 2013; Lu et al., 2015). Additional constraints can be applied based on double bond equivalent (DBE), which indicates the number of rings and double bonds in a molecule and is a measure of the degree of unsaturation in a given compound (Wozniak et al., 2008). Neutral molecules must have a DBE with an integer value (Koch et al., 2005). However, the elements N, S, and P may have different valences depending on their chemical environment so constraints based on DBE values need to be used with caution (Kind and Fiehn, 2007). Formulae are often filtered base on the "nitrogen rule" (Kujawinski and Behn, 2006; Kind and Fiehn, 2007; Koch et al., 2007). Neutral molecules containing an odd total number of ^{14}N atoms always exhibit an odd nominal mass (Koch et al., 2007). The nitrogen rule derives from the fact that chemical elements with even nominal mass have an even valence, while elements with odd mass have an odd valence, with the exception of nitrogen (Nizkorodov et al., 2011). The majority of data processing methods for Fourier transform ion cyclotron resonance technique (FT-ICR) and Orbitrap™ mass spectrometers check for the presence of isotopes rather than using isotopic ratios for formula assignments (Koch et al., 2007; Wozniak et al., 2008; Ohno et al., 2016). Other mass spectrometers, *e.g.* TOF-type, often use isotope patterns for compound identification which tend to provide more reliable assignments compared to FT-ICR and Orbitrap™ mass spectrometers (Glauser et al., 2013).

Once chemically meaningful formulae have been filtered, more than one possible formula per peak may still exist, especially at high *m/z*. In order to select the most meaningful formula assignment, two general strategies have been applied: a "best-fit" approach, in which the formula with the closest match between theoretical mass and observed

mass is selected, and a "formula extension" approach, in which chemical and structural relationships among compounds are taken into account for formula assignment, *e.g.* by looking for homologous series based on Kendrick mass defects (Hughey et al., 2001; Kujawinski and Behn, 2006; Putman et al., 2012; Danger et al., 2013). In the first case, possible incorrect assignments may arise from inaccuracies in the measured masses (Olsen et al., 2005; Reemtsma, 2009; Herzsprung et al., 2016). In the second case, it has previously been observed that, for example, atmospheric oxidation reactions involving S- and N-containing functional groups may lead to a wide variety of products that do not produce homologous series, risking the removal of potentially correct assignments (Kourtchev et al., 2016).

Most of the methods and currently available algorithms were developed based on ESI and therefore they rely on the assumption that ionisation is accompanied by protonation, deprotonation or adduct formation (Reemtsma, 2009; Kunenkov et al., 2009; Nguyen et al., 2010). Other ionisation techniques, such as atmospheric pressure chemical ionisation (APCI) and atmospheric pressure photoionisation (APPI) are becoming increasingly common for less polar and apolar organic compounds (de Hoffmann and Stroobant, 2007; Reemtsma, 2009). In APPI, detection of molecular ions (as radical cations or anions) over quasi-molecular ions is common (de Hoffmann and Stroobant, 2007; Reemtsma, 2009) so there is a need to develop new algorithms that take into account the formation of molecular ions.

Here we developed a code to filter molecular formula assignments that i) can be applied to different soft-ionisation techniques like ESI, and APPI in both positive and negative ionisation, ii) takes into account formation of molecular ions, quasi-molecular ions and Na adducts, iii) uses a novel method for mass shift and noise estimation and iv) can be used with two different blank subtraction methods. Many steps of the scheme are widely used in mass spectrometry studies, but direct comparisons between methods are difficult as detailed procedures are often not available in the literature. Aspects of the approach described here have been previously applied in studies of environmental (Kourtchev et al., 2013) and biological samples (Giorio et al., 2015).

2. Pre-processing

The following discussion is based on the use of an Orbitrap™ mass spectrometer (LTQ Orbitrap Velos, Thermo Scientific™, Bremen, Germany) with the proprietary software Xcalibur™ 2.1–3.0 (Thermo Scientific™, Bremen, Germany) henceforth referred to simply as *Xcalibur*. The steps taken, however, apply for the general processing of mass spectra with any spectrometer/software combination. While the choice of mass spectrometer does not affect the presented procedure, the resolution of the spectrometer will influence the accuracy of the final spectra with higher resolutions providing clearer peak separation and higher confidence in the molecular formula assignment (Schmitt-Kopplin et al., 2010) which is particularly important for complex ambient measurements (Mazzoleni et al., 2010).

2.1. Data acquisition

Ion transmission and ion collection efficiencies in an Orbitrap™ mass spectrometer strongly depend on the *m/z* scan range (Ranninger et al., 2016). Therefore, to avoid the loss of the ions at the high or low end of the selected mass range (*e.g. m/z* 50–1 000), it can be split into several overlapping scan ranges (Ranninger et al., 2016). Each of these scan ranges are independently processed and subsequently recombined. Both sample (*i.e.* the spectrum of interest) and blank (*i.e.* a reference spectrum) spectra are acquired under the same conditions. The type of blank will vary based on application but can include solvent, procedural, and field blanks. The blanks are later used to remove