

# Accurate mass measurements and their appropriate use for reliable analyte identification

A. Ruth Godfrey · A. Gareth Brenton

Received: 19 April 2012 / Revised: 15 May 2012 / Accepted: 21 May 2012 / Published online: 15 June 2012  
© Springer-Verlag 2012

**Abstract** Accurate mass instrumentation is becoming increasingly available to non-expert users. This data can be mis-used, particularly for analyte identification. Current best practice in assigning potential elemental formula for reliable analyte identification has been described with modern informatic approaches to analyte elucidation, including chemometric characterisation, data processing and searching using facilities such as the Chemical Abstracts Service (CAS) Registry and ChempSpider.

**Keywords** Accurate mass measurement · Mass accuracy and precision · Statistics for mass measurement · Fragmentation map/tree · High resolution · Database searching

## Introduction

Traditionally spectroscopic techniques, such as nuclear magnetic resonance (NMR), have been used for the elucidation of unknown analytes. However, these methods can be complex and expensive often with poor sensitivity. Liquid chromatography-mass spectrometry (LC-MS) is an analytical platform capable of rapid and high-throughput analysis with high sensitivity and is now commonly used in analyte identification. In addition to a retention time index provided by chromatography, mass spectrometry can offer further details of the chemical composition of the analyte in question by measuring its mass or mass-to-charge ratio ( $m/z$ ). The  $m/z$  of an ion can be described as either nominal (to the

nearest integer) or accurate. An accurate mass is ‘*the experimentally determined mass of an ion measured to an appropriate degree of accuracy and precision used to determine, or limit the possibilities for, the elemental formula of the ion*’ [1]. For reliable reporting of accurate mass data without ambiguity there is a requirement to statistically treat and use terminology that describes this data in a consistent fashion. The many publications involving accurate mass data have shown that this has not necessarily been the case with ‘error’ and ‘accuracy’ often used to describe mass deviation, where ‘error’ has been used to describe the deviation for a single measurement and ‘accuracy’ for a single and set of measurements. These variations in terminology for both statistical analysis and quoting of accurate mass data have been the subject of an earlier article [2] and will not be covered in detail herein. However, the authors will clarify that as stated in that tutorial,

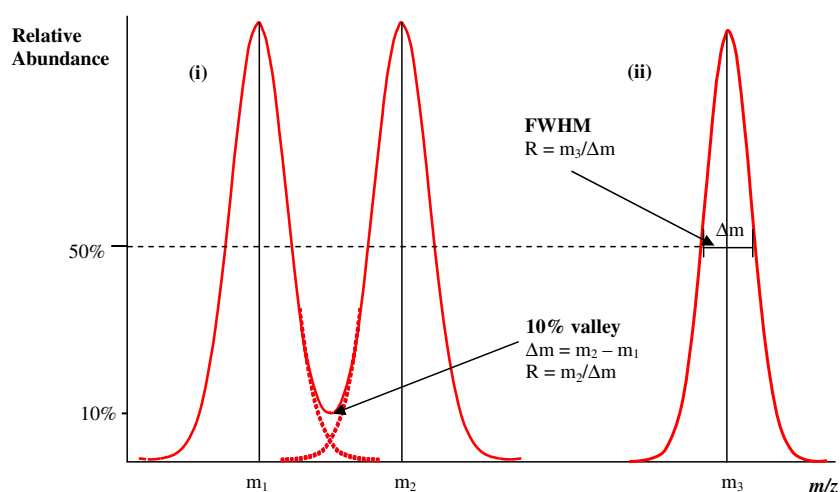
*‘the difference between the measured value (accurate mass) and the true value (exact mass) is the “accuracy” of the “accurate mass measurement” (an unfortunate double use of the word) and it is suggested that the term “mass measurement accuracy” should be used to denote this difference.’ [2]*

Therefore, both mass measurement accuracy and precision (repeatability) should be determined to characterise the natural variation in the acquisition of such data and provide a confidence limit with regards to the mass measurement accuracy for limiting potential elemental formula.

The ability to differentiate an analyte from another typically increases with increasing mass resolution [3]. Evidence of the claim in this initial statement is observed at a mass resolution greater than 10,000 at full width half maximum (FWHM) [3, 4], see Fig. 1 for the definition. Instruments with ‘highest-resolution’ mass analysers (resolving

A. R. Godfrey (✉) · A. G. Brenton  
Institute of Mass Spectrometry, Grove Building,  
College of Medicine, Swansea University,  
Swansea SA2 8PP, UK  
e-mail: A.R.Godfrey@swansea.ac.uk

**Fig. 1** Mass spectrum illustrating two definitions of resolution,  $R$ , (i.e. resolving power); (i) 10 % valley (ii) full width half maximum (FWHM). 10 % valley definition is based on the separation of two peaks of equal intensity with a valley at 10 % of each peak height. FWHM is based on the peak width at 50 % of the peak height



power at  $\text{FWHM} > 10,000$ ) that are appropriately calibrated are capable of separating neighbouring ions and acquiring accurate  $m/z$  data that may provide chemical information key to elucidating the elemental or molecular formula of an unknown analyte. For example, the subtle differences in exact masses of the elements and their isotopes from their nominal mass enables elemental formula to be derived from the measured  $m/z$ , if it is acquired to sufficient accuracy. However, it is unlikely that accurate  $m/z$  data only can be used for analyte differentiation (such as that obtained using lower resolution instrumentation with software manipulation [5]) as the number of possible elemental composition increases ‘exponentially’ with mass [6]. Hence, the elucidation of ‘unknown’ analytes will typically require the combination of both accurate mass and high resolution capabilities where the number of possible elemental formulae are further reduced using additional chemical information such as mass spectral peak purity (i.e. detecting isomers/isobaric species), an isotope pattern [3] and accurate mass differences (e.g. those observed from a neutral loss or formation of an adduct).

A mass spectrometer’s mass (i.e.  $m/z$ ) scale is established by connecting a physical property of the detected ion beam, such as time or frequency to mass. This process of mass calibration generally uses a mixture of known compounds, the calibrant, and is key to obtaining accurate mass data on a set of analytes. The analyst should be aware of the inherent limitations and accuracy associated with the alternative modes of calibration (internal, external, Lockspray™ etc. [3, 7, 8]). Using data integration (averaging) the effects of random error associated with the data acquisition can be reduced [2]. Without appropriate instrument calibration (including the  $m/z$  scale) the mass analyser may be inherently deficient in accuracy to which it can measure ions. An additional point of note to the analyst is the possibility of the presence of systematic errors during data acquisition. Unlike random error, a systematic error can result in mass

measurements that are different to the calculated exact mass, affecting accuracy and introducing bias in mass measurements. It is in the absence of systematic error that the ‘accuracy’ of accurate mass data should improve (tend toward zero) with the repetition of analyses (i.e. data averaging). Unfortunately, in practice most data acquisition is carried out with some degree of systematic error occurring from small instrumental instabilities or other effects such as ‘space-charging’ [3], resulting in a drift (offset) of the mass scale.

### Chemometric and qualitative analysis of accurate mass data

Statistical metrics that describe the mass measurement accuracy and precision, such as mass deviation or ‘error’ and standard deviation or root mean square (RMS) error respectively, can be used to indicate the presence and magnitude of systematic error (accuracy) or random error (precision) within the acquisition method. It is good practice to determine these metrics to characterise the uncertainty associated with mass measurement as this is dependent on the instrumentation, in particular, the mass analyser involved. Once established, the mass measurement accuracy and precision, along with the confidence limits of mass measurement are helpful in the intelligent selection of elemental formula for elucidating the identity of ‘unknown’ analytes (e.g. those within 3 times the standard deviation or standard error [5]). Combined with information associated with mass measurement described below these chemometrics enable this assignment process to be carried out with greater reliability.

#### Effect of instrumentation

Accurate mass determination has proven to be a key ingredient in analyte identification and has typically involved high

resolution (HR) mass spectrometry using mass analysers such as Fourier transform (FT), either Orbitrap or magnetic ion cyclotron resonance technology, reflectron time-of-flight (ToF) or magnetic sector technology. We have previously highlighted the effect of calibration method on mass measurement yet it is key that the analyst recognises that these instrument types have their own inherent accuracy and precision dependent on the design of the mass analyser. For example, FT-ICR and magnetic sector instruments are relatively higher resolution instruments, typically capable of sub-ppm mass measurement accuracy, while Orbitrap and ToF instruments are typically capable of accuracies greater than 1 ppm (see Table 1). Also, it is important to note that the method of mass measurement can differ between instrument types each with their own variations in mass measurement accuracy; data is typically acquired under full mass scan conditions using FT-ICR but using a ‘peak matching’ method with a magnetic sector [9] to achieve highest accuracy. Ultimately, the analyst should knowledgeably use an instrument and the accurate mass data generated, with consideration (and allowances) given to the factors that will affect mass measurement.

#### Pattern recognition: isotope ratios

Naturally occurring chemical species will contain a mixture of monoisotopic and isotopic masses dependent on their natural abundance. A mass spectrum of sufficient resolution can report the ionic species containing the monoisotopic elements (M) and an isotopic element (M + 1) or elements (M + n). Both the presence of isotopic peak(s) and their relative abundance to the monoisotopic peak can provide evidence for constraining the number of elemental formulae

**Table 1** Summary of typical performance characteristics (i.e. mass accuracy, mass resolution) of specific mass analysers [3, 21–26]. The reader should note that these are typical operating values and are obtained using various calibration techniques dependent on the type of mass analyser

Mass Analyser	Typical mass accuracy (ppm)	Typical mass resolution (FWHM) at $m/z$ 1000
Fourier Transform ion cyclotron resonance (FT-ICR)	1–2 <sup>a</sup>	10 <sup>5</sup> <sup>a</sup>
Magnetic sector	1–2	10 <sup>5</sup>
Time-of-flight (ToF)	~1–5 <sup>b</sup>	10 <sup>3</sup> –5×10 <sup>4</sup>
Orbitrap	1–2 <sup>c</sup>	10 <sup>5</sup>
Ion trap	~100	10 <sup>3</sup>
Quadrupole	~100	10 <sup>3</sup>

<sup>a</sup> Performance can greater exceed this typical operating value and greater mass resolution is achievable below  $m/z$ 1000 [3, 21]

<sup>b</sup> Certain manufacturers of ToF instruments do claim that sup-ppm accuracy can be achieved [25].

<sup>c</sup> Sub-ppm accuracy capable with software correction [26].

generated from an accurate mass measurement. Methods that compare the isotope patterns can vary; relatively simple methods measuring only the isotope ratio [10] or more complex methods measuring the accuracy of the continuous isotope profile known as a peak shape function [5]. The application of the latter method has proven to have additional advantages. The method itself and the limited data loss from using a profile provides a two dimensional measure of uncertainty of instrument response (relative abundance) across the mass scale. This can be used to improve the mass measurement accuracy of lower resolution instrumentation, such as a quadrupole [5], in addition to providing an accurate representation of the isotope profile for an unknown analyte for comparison to a suspected elemental formula.

#### Heuristic rules of elemental formula assignment using accurate mass data

This article has thus far covered essential components for characterising the uncertainty of accurate mass measurement in order for this data to be used reliably with confidence. It is well accepted that a number of elemental formulae may be valid for such an accurate mass measurement and prior to assigning a chemical structure to an unknown analyte an elemental formula will need to be selected to limit the number of possible chemical structures. A knowledgeable method for implementing such a filter is to apply chemical and heuristic rules for elemental formulae described within Table 2 primarily sourced from Kind *et al.* [10]. These include relatively well known chemical rules such as the nitrogen and ring-plus-double-bond equivalence rules [11], and established principles relating to valence electron theory (i.e. Lewis ‘octet rule’) [10]. The analyst should note that the individual use of these rules is not without problems [10] and evidence for a particular elemental formula generated from them is best treated collectively. However, the application of all these rules jointly should enable a sensible elemental formula to be selected to minimise computational resource in providing a chemical structure.

#### Use of accurate mass data for analyte identification

##### Accurate mass fragmentation (MS<sup>n</sup>) maps

Analyte (and chemical structure) identification may be possible using first principles, elucidating structure according to the fragmentation pattern of the analyte [12–14]. Early methods typically used electron ionisation (EI) and produced large amounts of fragmentation data during the ionisation process. Therefore only MS and MS/MS instrumentation were necessary to determine the order of fragmentation, integral to

**Table 2** Summary of elemental formula filter rules sourced from the literature [10, 11, 18, 20]

Name of rule	Details
Rings-and-double-bond equivalence	$RDBE = C + Si - \frac{1}{2}(H + F + Cl + Br + I) + \frac{1}{2}(N + P) + 1^a$
Nitrogen	Integer mass = odd, #N = odd <sup>b</sup>
Restriction of element numbers	Use the absolute element number for the $m/z$ (i.e. if measuring $m/z$ of 850, it will contain $< (850/12) = 71$ carbon atoms.
LEWIS	Measures the occupancy of s and p orbitals <sup>c</sup> based on Lewis 'octet rule', can highlight presence of OE species.
SENIOR	Measures the occupancy of s and p orbitals <sup>c</sup> but takes account of additional criteria to LEWIS: 1. Total valences or number of atoms with odd valences is even. 2. Total valences $\geq 2 \times$ maximum valence. 3. Total valences $\geq 2 \times$ (number of atoms-1).
Isotope pattern	Measure presence and abundance ratio of isotopes.
Hydrogen/carbon element ratio check	Typically $0.2 < H/C < 3.1$ , but can be $0.1 < H/C < 6.0^d$
Heteroatom (N, O, P, S)/carbon element ratio check	Typically $< 1.3$ , but can be used to identify specific classes if low or 0 (i.e. alkanes/alkenes etc.).
Multiple element probability	Minimises formula with certain combinations of high numbers of heteroatoms.
Derivatisation adduct (e.g. trimethylsilyl TMS) check	Subtract formula of derivatisation adduct from observed elemental formula if used as can obscure information generated from other filter rules.
Reference database search	Use database such as CAS or Chemspider and search according to molecular weight/formula or monoisotopic mass ( $m/z$ )

<sup>a</sup> can provide information whether even electron (EE) or odd electron (OE) ion. Typically electrospray ionisation (ESI) produces EE ions and thus formulae can be filtered according to what is expected with ionisation technique

<sup>b</sup> applicable for masses  $< 500$  Da

<sup>c</sup> for neutral species only, i.e. masses must be determined without the proton adduct for ESI

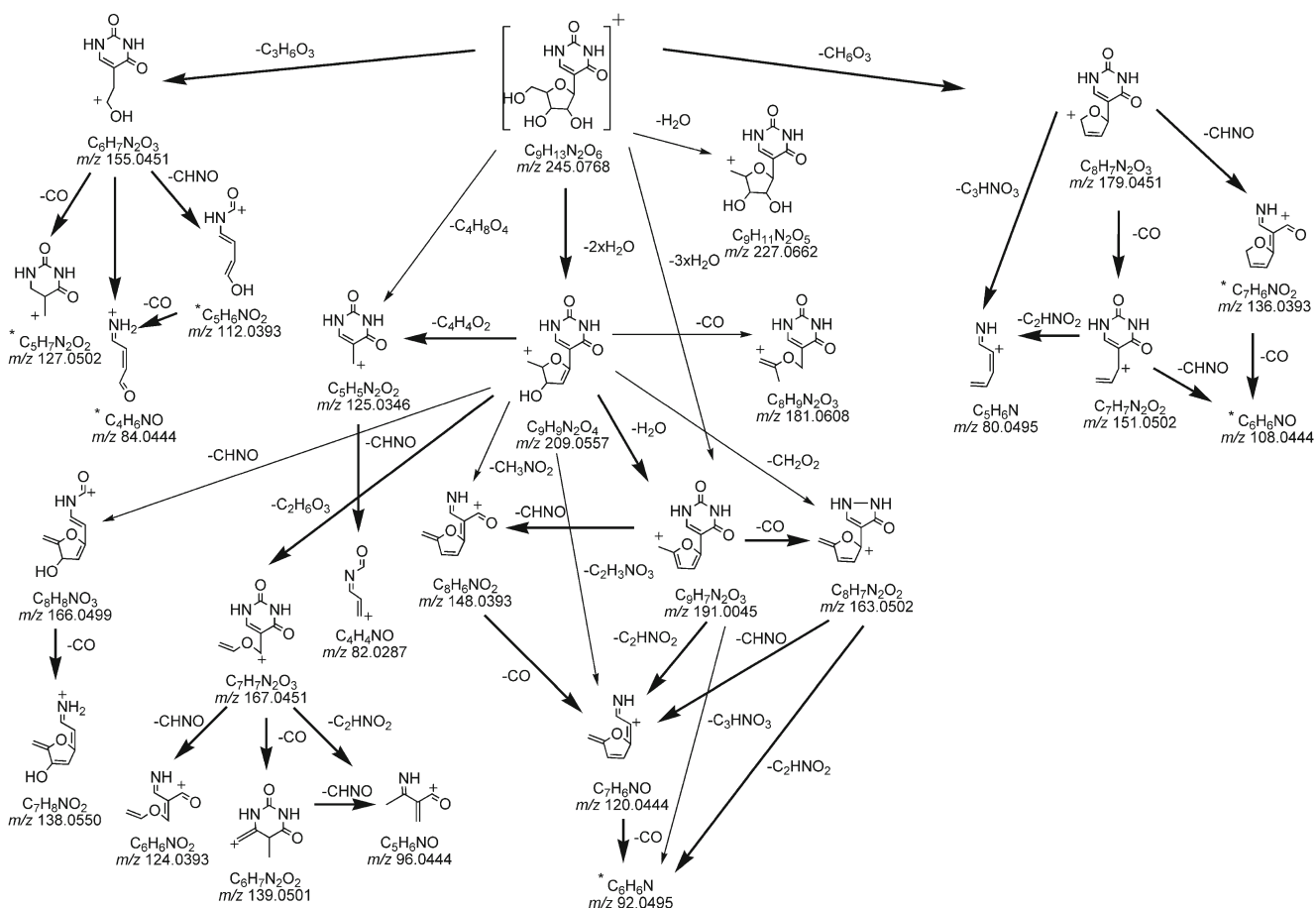
<sup>d</sup> extremes out of 'common range' can be assigned to specific chemical classes (e.g. fullerenes  $\approx 0.15$ )

assigning the position of chemical groups within the structure. More modern 'softer' ionisation techniques such as electrospray ionisation tend to produce very little fragmentation during the ionisation process and often will require multiple,  $MS^n$ , stages of fragmentation using an ion trap mass spectrometer to generate sufficient detail. The sample throughput of this method can be dramatically increased for unknown analytes by using a data-dependent acquisition (DDA) where ions are chosen for fragmentation when they appear within a mass spectrum [13]. Combined with high mass resolution capabilities the analyst may provide a fragmentation map with elemental formulae for the ions present, confirming the calculated neutral loss between successive stages (see Fig. 2). The interpretation of such fragmentation mechanisms can be aided by using software packages such as Mass Frontier (Thermo Fisher Scientific, San Jose, CA) or ACD/MS Fragmenter (ACD/Labs, Toronto). An advantage of using such software is that the raw data and the proposed structural information may be stored (in the form of a fragmentation tree [14]) as a database of analytes with the potential to be used for analyte identification in further studies. This approach may offer an efficient method for analyte identification (similar to that employed with the EI NIST database [15]) using soft ionisation techniques and without the need for obtaining a reference standard for confirmation.

## Chemical databases

Computer assisted structure elucidation (CASE) has been a key area of development in the last 30 years [16, 17]. An effective method for analyte identification uses accurate mass data that is searched in a 'spectraless' chemical database such as Chemical Abstracts Service (CAS) registry or Chemspider [18–20]. Searches can be carried out using a range of properties but typically use elemental formula or molecular weight as the active search parameter where results are ranked according to the highest number of references/hits within the database. The use of databases should not discourage employing other data commonly generated when using mass spectrometry (and related techniques). For example, chromatographic retention time, fragmentation and isotope patterns can prove essential in distinguishing between analytes with similar mass (and therefore elemental formula) properties.

It has been recently reported that the Chemspider interface has been modified to further accommodate mass spectrometry data where users can search the database according to the monoisotopic accurate mass [20], unlike a CAS search [18]. This is considered to have a lower uncertainty associated with the search versus molecular weight and does not assume that a possible elemental formula chosen for the search is correct. Miscalculations can often occur with



**Fig. 2** Accurate mass fragmentation map of pseudouridine taken from Godfrey *et al.* [13]. In this illustration the large headed arrows represent the most abundant ions in the relevant fragmentation spectra indicating the most preferable fragmentation paths for the structure. This map was generated as part of an online LC-MS<sup>n</sup> method using a

data dependent acquisition (DDA) approach and collision induced dissociation (CID) method of fragmentation. Data was generated as repeat acquisitions and structures elucidated with the use of Mass Frontier (Thermo Fisher Scientific, San Jose, CA)

determining and searching according to the molecular weight (e.g. in error of the electron for protonated species) as the mass spectrometer operating software typically reports the monoisotopic mass and the molecular weight must be determined by the analyst. It is understood that a search by monoisotopic mass has proven to be as effective, if not more, when compared to elemental formula under certain conditions (molecular weight >600 Da [20]) and may provide an unbiased search from the original raw data. In light of this evidence and under future circumstances where the number of Chemspider entries surpasses that of the CAS Registry, this modified interface may prove key in establishing Chemspider as the gold standard database for identifying unknown analytes.

## Outlook

Accurate mass data obtained with high resolution mass spectrometers can be an effective tool for analyte identification. There are a number of approaches, both manual and informatic,

that may be used for improving the accuracy and precision of accurate mass data which is essential for us as analysts to understand the limitations, validity and usability of such data. This is increasingly more important in modern analytical laboratories as the amount of sample for applications typically encountered are becoming less and less. Hence, the requirement for a 'one pass analysis' is becoming a more common scenario. The approaches described within this manuscript are intended to achieve this aim and when combined have the potential to offer a powerful analytical tool for analyte identification. A potential workflow could involve high resolution accurate mass data acquired using a DDA method generating accurate mass full scan and MS<sup>n</sup> data that is uploaded into in-house database software such as Mass Frontier. A monoisotopic accurate mass and an isotope pattern from data post-processed with the spectral accuracy function [5] may be used to reduce the possible number of elemental formulae, thus creating a more accurate, accurate mass data set for database searching. Using the chemical and heuristic rules described by Kind *et al.* [10], these formulae may be filtered further, possibly

to a single elemental formula for lower mass species. This data set could be searched against an in-house database or a larger database, such as Chemspider, using either the monoisotopic mass or elemental formula function for a potential chemical structure. For added confidence, both searches may be carried out for confirmation of the result or with an alternative database such as the CAS registry.

There are a number of strategies available to the modern mass spectrometrists to enhance the informatic usability of accurate mass data. The judicious use and combination of these 'tools' may offer a powerful method for not only the identification of unknown analytes but also maximising the informatic content of the data acquired with appropriate storage for future use. This in conjunction with open access rights of chemical databases like Chemspider has the potential to rapidly increase the volume of reliable data available for expert and novice mass spectrometry users alike, and play an integral role in supporting the needs of industry, research and education.

**Acknowledgments** Contract/grant sponsor, Engineering and Physical Sciences Research Council (EPSRC); contract/grant number, EP/F014341/1.

## References

1. Gross ML (1994) *J Am Soc Mass Spectrom* 5:57
2. Brenton AG, Godfrey AR (2010) *J Am Soc Mass Spectrom* 21:1821–1835
3. Marshall AG, Hendrickson CL (2008) *Annu Rev Anal Chem* 1:579–599
4. IUPAC Gold Book. Available online <http://goldbook.iupac.org/>. Accessed March 2012.
5. Wang Y, Gu M (2010) *Anal Chem* 82:7055–7062
6. Kim S, Rodgers RP, Marshall AG (2006) *Int J Mass Spectrom* 251:260–265
7. Eckers C, Wolff JC, Haskins NJ, Sage AB, Giles K, Bateman R (2000) *Anal Chem* 72:3683–3688
8. Wolff JC, Eckers C, Sage AB, Giles K, Bateman R (2001) *Anal Chem* 73:2605–2612
9. D'Agostino PA, Hancock JR, Provost LR, Semchuk PD, Hodges RS (1995) *Rapid Commun Mass Spectrom* 9:597–603
10. Kind T, Fiehn O (2007) *BMC Bioinformatics* 8:105
11. McLafferty FW, Turecek F (1993) University Science Books, CA.
12. Beynon JH, Morgan RP, Brenton AG (1979) *Phil Trans Roy Soc A293*:157–166
13. Godfrey AR, Williams CM, Dudley E, Newton RP, Willshaw P, Mikhail A, Bastin L, Brenton AG (2009) *Rapid Commun Mass Spectrom* 23:3194–3204
14. Rasche F, Svatos A, Maddula RK, Bottcher C, Bocker S (2011) *Anal Chem* 83:1243–1251
15. NIST 11 database. Available online <http://www.sisweb.com/software/ms/nist.htm>.
16. Carhart RE, Smith DH, Brown H, Djerassi C (1975) *J Am Chem Soc* 97:5755–5762
17. Buchanan BG, Smith DH, White WC, Gritter RJ, Feigenbaum EA, Lederberg J, Djerassi C (1976) *J Am Chem Soc* 98:6168–6178
18. Little JL, Cleven CD, Brown SD (2011) *J Am Soc Mass Spectrom* 22:348–359
19. Wolff JC, Thomson LA, Eckers C (2003) *Rapid Commun Mass Spectrom* 17:215–221
20. Little JL, Williams AJ, Pshenichnov A, Tkachenko V (2012) *J Am Soc Mass Spectrom* 23:179–185
21. Marshall AG, Hendrickson CL, Jackson GS (1998) *Mass Spec Rev* 17:1–35
22. Bristow AWT, Webb KS (2003) *J Am Soc Mass Spectrom* 14:1086–1098
23. Hardman M, Makarov AA (2003) *Anal Chem* 75:1699–1705
24. Makarov A, Denisov E, Lange O (2009) *J Am Soc Mass Spectrom* 20:1391–1396
25. Stroh JG, Petucci CJ, Brecker SJ, Huang N, Lau JM (2007) *J Am Soc Mass Spectrom* 18:1612–1616
26. Cox J, Mann M (2009) *J Am Soc Mass Spectrom* 20:1477–1485