Mass assignment, centroiding, and resolution – fundamental definitions and confusions in mass spectrometry

Jan Urban[1,2,*], Nils Kristian Afseth[1], and Dalibor Štys[2]

1-Nofima – Norwegian Institute of Food, Fisheries and Aquaculture Research, Osloveien 1, 1430 Ås, Norway.
2-University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, Laboratory of Applied System Biology, Zámek 136, Nové Hrady 373 33, Czech Republic.
**-Corresponding author, urbanj@frov.jcu.cz, phone +420 38 777 3842, Zámek 136, Nové Hrady 373 33, Czech Republic.*

Abstract

Even though the main steps of preprocessing and data analysis in Liquid / Gas Chromatography-Mass Spectrometry have been frequently reviewed in recent years, little attention has been put on the initial processing of these data, from mass detection and centroiding to the use of the fundamental definitions such as resolution. The choices made in this initial part of analysis will severely affect the end result of the analysis, and this article presents a current approach to the decomposition of the mass spectrum into mass peaks and the estimation of mass centroid positions. In addition, recommendations on the use of fundamental definitions are often confusing and inconsistent across the literature. Although this conflict in terminology has been reported, different definitions are still supported. Thus, in this paper, recommendations and analogies are discussed. Topological terms of distinguishability and discriminability are also introduced to discern between the theoretical ability of a detector to distinguish adjacent MS peaks, and what could actually be achieved.

Keywords: Liquid chromatography, gas chromatography, mass spectrometry, resolution, resolving power, distinguishability, discriminability, mass peak, FWHM, Savitzky-Golay.

1. Introduction

Knowledge of the characteristics of the measurement process itself is crucial in any processing and analysis of measured data. The ignorance of these characteristics and their uncertainties will be propagated through the processing computations and may subsequently lead to an incorrect interpretation [1,2]. One of the most deterrent examples in history was the destruction of NASA's spacecraft "Mars Climate Orbiter" during orbit insertion because of unit's mismatch [3]. The mathematical abstraction of any characteristic is described via attributes. Unfortunately, some of the attributes are often improperly interchanged even in recommendations and regulatory documents [4,5]. Thus, it is of utmost importance to define the exact meaning of each term before use.

Liquid (LC) or gas (GC) chromatography in tandem with mass spectrometry (MS) is widely used in many chemical and biochemical analytical setups, especially in the so-called "-omics" sciences, and the techniques are used as key tools for unraveling biochemical pathways within systems biology [6-8]. The preprocessing of these types of data, including feature detection, alignment, and normalization, with subsequent multivariate data analysis, is an essential part of understanding and interpreting the results. The output from LC/GC-MS, and thus the input for data processing and data analysis, might be separated into three distinct groups of attributes:

I) Attributes of the obtained data, e.g. retention time, mass-to-charge ratio, intensity (counts), and derived attributes like TIC (Total ion current chromatogram), number of time scans,

maximal intensity, and mass range.

II) Attributes of the measured sample, like the sample origin or a description of the preparation procedure. These are the attributes usually relevant only for analysis and interpretation, not for the data processing. (9).

III) Attributes of the measurement device and its abilities. These are the fundamental attributes from the theory of measurement.[10]

In LC/GC-MS, the retention or elution time (*rt*) is the main attribute in the chromatographic domain, while the mass-to-charge ratio (*m/z*, or *MZ)* is the main attribute in the mass spectrometry domain. Whereas the units of the former attribute are unambiguously defined, the mass-to-charge variable is not. This variable is frequently expressed in unified atomic mass units, recommended by IUPAC [11, 12], but Dalton and Thomson units (*Th*, [13]) are also used. The unified atomic mass unit and the Dalton unit are not part of the SI system, but it is recognized by CGPM [14] that they will continue to be used in appropriate contexts [12]. *Th* is not an SI unit, and it has not been accepted by IUPAC. Discussions are still ongoing in the MS community on which units that should be used [15-20].

Resolution and resolving power are other frequently used terms in LC/GC-MS analysis. These terms, however, are very often interchanged, and even though the conflict in terminology has been already reported [4,5], different definitions are still supported [21-27]. A similar challenge within optical resolution was reviewed by den Dekker and van den Bos [28]: "*In applied science, resolution has always been, and still is, an important issue. Since, it is not unambiguously defined, it is interpreted in many ways.*" Thus, this confusion is not specific just for the mass spectrometry field, but has a generic perspective.

Even though the main steps of preprocessing in LC/GC-MS analysis have been frequently reviewed in recent years [2,4,9,10,15,18,23,24], little attention has been put on the initial processing of these data, from mass detection and centroiding to decisions related to mass resolution. The choices made in this initial part of analysis will severely affect the end result of the analysis, and in this paper, thus, an overview and summary of definitions and recommendations for the initial part of preprocessing in LC/GC-MS is provided. Section 2 describes the mathematical approaches that are widely used for LC/GC-MS data pre-processing. Different criteria for practical evaluation of such parameters from the theoretical recommendations/definitions, as well as the definitions themselves, are discussed in section 3. In section 4, we introduce the formal definitions of the fundamental attributes in mass spectrometry from a topological space point of view to advocate some of the recommendations from section 3.

2. Mass assignment and centroiding

A typical dataset from LC/GC-MS measurements is represented as a discrete set of points in a discrete three dimensional space which is defined by discrete axes, namely retention time (*rt*), mass to charge ratio (*m/z*), and intensity (counts), as shown in figure 1. Analytes (components of the system being analyzed) elute at specific time points from the chromatographic column and enter the MS ionization chamber. The time delay of elution of the different analytes is caused by physicochemical interactions between the stationary phase on the chromatographic column, the analytes and the mobile phase. The intensity at each detectable *m/z* is registered inside the MS detector and its value represents the approximate amount of detected ionized molecules of each individual *m/z* at the exact retention time. In practice, the continuous signal is sampled using an analog-to-digital converter, a non-ideal device with various physical limitations. All real signals are discretized, quantized, and reduced into a discrete finite set of values. Discrete-valued signals are always just an approximation to the original continuous-valued signal [29].

One 'slice' of the 3D data set selected at one specific *rt* is a mass spectrum, as an accumulation of all detections on the MS detector during a very short time period. The detector requires some time interval to provide the detection as well as to recover for the next counting. The mass spectrum represents a discretized distribution of ions by the mass to charge ratio.

Unfortunately, even the beam of ionized molecules at the same *m/z* value contains ions with different vectors of kinetic energy. Therefore, the trajectories of the ions are distributed in some width of the beam and will contribute to broadening of the mass peak on the detector [4].

The mass spectrometer can record mass spectra in two different modes: the profile mode and the centroid mode. While the profile (or quasi-continuous) mode preserves the shape of the mass peak (within limits of discretization), the centroid mode records only a weighted average of the mass peak. This is shown in figure 2 for different types of mass spectrometers. Ideally, when symmetrical and smooth peak shapes are produced, centroid positions are equivalent to the positions of the local maxima, and the peak borders are equivalent to the local minima. However, if there are any contributions from adjacent or overlapping components, the peak shape is distorted [30, 31]. Magnetic sector instruments produce triangular or Gaussian centroid curves, while shapes from quadrupole analyzers are trapezoid or flat-topped. Ion traps and time-of-flight detectors tend to give centroid peaks with sharper apexes or with increased widths at the base [30]. The mass domain in cyclotron resonance and orbitrap mass spectrometers results after inverse Fourier transformations of the frequency domain, thus the mass peak shapes are represented by sinc (cardinal sine) shapes with Cauchy-Lorentzian envelopes [32].

As all subsequent data analysis of LC/GC-MS data, from noise reduction to feature extraction, presumes centroided values, the proper conversion from profile to centroid mode is a fundamental issue in pre-processing. The computation of the centroid position *C* on the *m/z* axis is defined in equation 1:

$$C = \sum_a^b (y \cdot m) / \sum_a^b (y), \quad (1)$$

where *m* represents the *m/z* position, *y* represents the intensity, $\sum_a^b (y)$ represents the area of the mass peak, and *a and b* represents the borders of the mass peak.

Thus, in order to calculate accurate centroid values, the peak position, the peak shape, as well as the peak borders have to be properly determined. There are several basic approaches available to determine these features. One way is to fit a proper distribution function according to the individual ion beam. The most common approach is to create a class of possible shape models (e.g. triangular or Gaussian models) and choose the most appropriate. However, even when understanding the underlying physical and chemical properties of the MS instrument, it is a non-trivial task to select the proper shape model. Hence, regression analysis is frequently used [33] to find the model using the minimal error criterion between detected and modeled peak shapes. Unfortunately, any evaluation of the parameters used can only help to decide which shape of the considered shapes is 'optimal' [34]. Very strong fit still does not mean that the best distribution function was considered [35].

Two parameters are necessary to fit a symmetric shape distribution: the location and the scale parameters of the fitting function. The location parameter refers to the position (apex) of the maximal or mode value of the mass peak (which ideally already is the centroid position). The scale parameter represents a measure of the spread of the distribution (or standard deviation). In other words, the scale parameter is related to the width of the beam as well as to the *m/z* interval delimited by the peak borders *a* and *b*. In figure 3, peak shape fitting using Gaussian and Laplacian distribution fitting, respectively, is illustrated.

Asymmetric (skewed) and noisy peaks require smoothing or reshaping. Usually, these transformations are obtained by spline or wavelet functions. In practice, such functions are always discrete apodization filters with specified window lengths that are applied piecewise along the axis of the original signal. The window of the filter iteratively moves along the signal's first dimension (which is the *m/z* axis for mass spectra), subsequently producing a transformed or centroided signal.

Numerous wavelet filters have been introduced during the last century. Some typical wavelet filters include the Bartlett (triangular) filter, the Hamming filter, the Gaussian filter, the Blackman filter, and the Ricker filter (frequently denoted Mexican hat). The crucial issue of the wavelet approach is to select an appropriate filter function and subsequently decide the length(s) of the window. This is illustrated in figure 4. The window length of the wavelet function is comparable with the scaling parameter of the fitting function, related to the mass peak width defined by the *m/z*

position and the peak borders (*a* and *b* in equation 1). The output of a wavelet transformation can be the composition of many window functions instead of only one, as shown in figure 5. If the shape of the mass peaks is unknown, spline transformations are often better than wavelet transformations. The spline transformation is a polynomial function approximation that is able to approximate an a priori unknown shape by a polynomial function of a low degree. While the lower degree preserves the raw shape very well, peak smoothing is likely to happen spontaneously. On the other hand, at higher degrees the approximated function might produce artificial signal oscillations on the peak borders (Runge's phenomenon). Spline transformations provide piecewise polynomial approximations, meaning that the whole peak is decomposed into short intervals. Each interval is then fitted by its own polynomial function. Therefore, the total shape fit consists of multiple polynomial pieces. While the interpolation is excellent inside the peak, extrapolation produces unusable values, such as oscillations at the peak borders.

Spline transformation preserves the positions of the peak maxima and minima. The most used and well cited is the Savitzky-Golay filter [38]. The Savitzky-Golay filter provides an approximation of the underlying peak shape by averaging polynomial windows of higher orders (usually 4^{th} degree polynomials). This approach was originally developed for spectroscopic applications in the time domain, and the approach is almost unknown in other scientific areas where filtrations, fittings, approximations or interpolations are required [37]. As was pointed out by Persson and Strang, "*it is not a tremendously powerful filter, but it's virtues are simplicity and speed*" [39].

As discussed in this chapter, distribution fitting, wavelet or spline transforms are different tools to estimate the mass peak centroid and its border positions. The performance of each method is strongly dependent on the input scaling parameter or the window lengths, which thus directly corresponds to the mass peak width. Hence, the decision to use any of these techniques might not be as essential as knowing the distance between two valid maxima or minima on the mass axis [40]. Therefore, the proper centroid position value *C* arises from the mass peak width definition, which corresponds to the determination of resolution and resolving power.

3. Resolution versus resolving power

The IUPAC GoldBook [36] offers three approximately equivalent descriptions of the term "resolution" (R05318) in mass spectrometry and two descriptions for the terms "resolving power" and "mass resolving power" (R05321, M03730). The recommendations are often cited, described, or interpreted [21-24, 41]. Let us have a look at the expressions and examine their meaning and consequences (Tables 1 and 2).

The first description of resolution (R05318) refers only to the resolution energy as a value derived from a peak showing a number of ions (intensity given in counts) as a function of their translational energy [23, 36, 41].Next, the *valley definition* states that if two mass spectrum peaks at masses $m - \Delta m$ and m of **equal heights** are separated by a valley which at its lowest point is equal to 10% of the height of the peaks [30, 49, 55], then the resolution is provided by equation 2:

$$R = m / \Delta m \quad (2)$$

Here the resolution R is a function of *m.* Thus, it is not a constant value across the dynamic range of the m/z axis, a fact that is often overlooked [40]. The value of the ratio $m / \Delta m$ represents an interesting *theoretical* property: if each peak has a width that equals Δm (or more precisely, if the distance between each two consecutive valid peak maxima equals Δm), then on the range between *0* and *m* there could be exactly R distinguishable peaks. However, Δm might vary according to *m.*

The *peak width definition* for a single peak expresses Δm as the width of the peak at a height which is a specified fraction of the maximum peak height (50%, 5%, or 0.5% is recommended). It is important to pinpoint that Δm is not the peak width. More precisely, Δm is the peak width at a given fraction of the maximum, and the used fraction should always be specified. According to the valley

definition, Δm is the difference between m/z positions of two maxima for which the 10% valley value is fulfilled. The valley definition of Δm is "technically equivalent" [36, 41] to the peak width at 5% of the peak height, if and only if the peak is isolated and symmetrical, and that linearity is guaranteed between the 5% and 10% levels of the peak height. However, isolated peaks are not frequently encountered in real mass spectra; symmetry is often distorted by random noise contributions. The linearity condition refers to the linearity [42] of the sensor response [29, 43]. Then the "peak width" used in the valley definition is equal to the peak width at 5% of the height of an isolated symmetrical peak.

The IUPAC peak width definition states a common standard as the 50% fraction, the Full Width at Half Maximum (FWHM), which is sometimes also improperly denoted "half width" [36]. The relation between the FWHM and the scale parameter of an ideal Gaussian peak is given in equation 3:

$$\mathrm{FWHM} = \sigma\, 2\, \sqrt{(\, 2 \ln(2)\,)}, \quad (3)$$

where σ is the standard deviation. Examples of relations for typical peak shapes are provided in Table 3. But why is it important to consider which peak fraction is the most meaningful? The valley definition of mass resolution is contingent upon two adjacent, mass peaks of equal size and shape, which is almost never the case experimentally [44]. The peak width definitions, on the other hand, are uncertain. Each fraction becomes valid only **after** the approximation: we want the Δm, which can be obtained from the approximated shape. However to approximate the shape, we need to know the value of Δm (or scale parameter or window length) as the input parameter.

Six simple examples on artificially created discrete mass peaks are illustrated in figure 6. The examples interpret the IUPAC definitions on resolution together with several important practical consequences. The first example (figures 6a-d) compares the valley and the 5% peak width definition for peaks of equal heights. The second and the third example (figures 6e-h) extend the definitions to peaks of different heights. Finally, the fourth, the fifth, and the sixth example (figures 6i-l) compare the valley definition with the FWHM.

A single isolated symmetrical Gaussian peak *m* with its centroid at *m/z = 102* is plotted (solid line) in the beginning of the first example (figure 6a). The 5% fraction (dashed line) of the peak height provides a peak width value of exactly 1 *m/z* unit. Then the resolution (R) equals R = 102 / 1, according to equation 2. The 50% fraction (FWHM, dotted line) could then also be evaluated.

In figure 6b another symmetrical Gaussian peak *m* with centroid at *m/z = 101* is added (dotted line). The new peak has exactly the same height, shape, and therefore also the same scale parameters and peak width at the 5% fraction (dashed line) as the former peak. The two peaks intersect exactly at the position of the 5% fraction level. The presence of the two mass peaks will be detected as a superimposed signal (solid line), as shown in figure 6c. The 5% fraction (dashed line) of the isolated peak is now **below** the valley between the two peaks.

However, as shown in figure 6d, the value of the valley of the superimposed signal (solid line) at its lower point is exactly at the 10% fraction (dashed line) of the peak height. While the distance between the centroids equals the 5% peak width (102 – 101 = 1), the 5% peak width of the isolated peak and the distance fulfilling the 10% valley are equivalent. This is the reason why the 5% fraction of an isolated symmetrical peak is technically equivalent to the valley definition.

The situation of two adjacent peaks with equal height is not usual in real experiments [44]. The second example (figure 6e) starts with an isolated symmetrical mass peak with the same centroid and scale parameter as the peak from figure 6a. However, the height of the peak is only half of the former peak.

Then another peak, with similar characteristics as in figure 6b, is added with its centroid at *m/z = 101* (figure 6f).

The superimposed signal (figure 6g) is detected by the MS detector. The valley of the superimposed signal between the peaks is below the 10% fraction of the highest peak, and it is also

above the 10% fraction of the lower peak. Therefore, the valley should be considered valid and the distance between the peak maxima is again equal to the 5% peak width.

What then if the height of one of the peak is much lower than the height of the second peak? The third example (figure 6h) illustrates a modified situation of figure 6g. The mass peak with a centroid at *m/z = 102* is exactly 20 times lower than the mass peak with its centroid at *m/z = 101*. The scale parameters and the positions remain unchanged, and the *102* mass peak height is even below the 10% fraction of the *101* mass peak. However, there is still an observable valley, and again, as shown in figure 6g, the valley is below the 10% fraction of the higher peak, and above the 10% fraction of the lower peak. Therefore, the valley should be considered valid.

This means that the 10% valley definition could be extended for any two adjacent peaks, without the condition of equal height. If the valley between the peaks is above the 10% of the lower peak and simultaneously the valley is below the 10% of the higher peak, then the distance Δm between peak maxima approximately equals the peak width at the 5 % fraction of the isolated peak. The order of peaks is not relevant, but the value of the IUPAC resolution always has to be computed from the peak of the higher m/z position.

The fourth example (figure 6i) illustrates a situation where the second peak is half the height of the first peak. The *FWHM* now has the same value as Δm computed via the valley definition in all previous examples. The resolution R (as provided in equation 3), computed using the FWHM, produces the same value as the resolution using the 10% valley of figure 6d. Numerically, the same resolution is obtained.

Graphically, however, the interpretation is somewhat different. The superimposed signal on the MS detector has no valley, as shown in figure 6j. In other words, the two peaks are not distinguishable. The individual peaks have bigger scale parameters as they produce the FWHM of the same value as the valley or 5% Δm in figure 6d. Obviously, the meaning of 5% Δm and 50 % Δm are not analogous. Actually, the 5% Δm is always 2.0792 times FWHM for Gaussian shapes. Moreover, the resolution values computed from differently defined Δm always have to be interpreted differently. The details will follow in the two subsequent examples.

The fifth example (figure 6k) introduces a case where the valley between the peaks becomes observable in the superimposed signal. If the scale parameter will be just a little bit higher, the valley disappears. The FWHM values of both peaks could be easily estimated. While there are two distinct apexes, the half width of the half maximum could be computed and then multiplied by two. However, the peaks are so close that the maximum of the superimposed signal are not on the same *m/z* position as the maxima of the individual peaks, but slightly shifted towards each other. The estimated FWHM values are *0.8142 m/z* for the lower peak and *0.9590 m/z* for the higher peak, whereas the FWHM of the lower individual peak was 0.8242.

It is thus important to realize that the peak width definition is describing isolated peaks, which is clearly not the case here. Moreover, the IUPAC peak definition states that the resolution **may be** expressed as $m / \Delta m$. It tells nothing about how to distinguish or separate these two peaks. The instruction of 'separation' is in the valley definition, but the 10% condition is far to be fulfilled in this example. What if the valley is not a real valley, but the product of the noise contribution? Unfortunately, none of the recommendations is helping in the decision here.

The last example (figure 6l) illustrates the worst case scenario where two individual peaks of equal height intersect or overlap. Both peaks are sufficiently broad that the superimposed signal produces a new (false) maximum instead of a valley between the two maxima. In this case it is not possible to distinguish if the detected signal is the superimposition of two peaks or if it is an individual peak. The FWHM of the superimposed peak was calculated to 2.0174 *m/z*.

Generally, the FWHM is used to describe the measurement of a peak width when that peak does not have sharp edges. On the other hand, the scale parameter does not describe the total width of the profile, as it theoretically extends forever [45]. The width across the profile when it drops to half of its maximum is a simple and well-defined number which can be used to compare the measurements obtained under different conditions. The only problematic issue of the FWHM is in the computation of the resolution. The $R = m / (50\% \Delta m)$ is approximately double than the

R = $m / (5\% \Delta m)$. The difference is this large because the 5% Δm and $50\% \Delta m$ represents different characteristics of the measurement. While the 5% Δm is equivalent to the distance between the maxima, the interpretation of the 50% Δm is not that simple.

It is common to express any variable (and in this case m) as the mean ± z proportions of standard deviation. The interval expressed as the FWHM ($2z = 2.355\sigma$) covers only 84.27% of the peak values. However, if the FWHM is considered as the $\pm \Delta m$ concept, then the interval of $2\,z = 2\,FWHM = 4.71\sigma$ covers 99.53% of the peak. Thus, the analogy of the 5% and 50% fraction peak widths could be approximated as $5\% \Delta m = 2\,(50\% \Delta m)$. The IUPAC resolution (R05318) is then defined as in equation 4:

$$R = m / (5\% \Delta m) = m / (2\,FWHM) \quad (4)$$

The term "resolving power" describes the ability to distinguish between ions differing in the quotient of mass/charge by a small increment. The IUPAC description of resolving power in mass spectrometry (R05321) states that the resolving power might be characterized by the peak width at 50% and at 5% of the maximum peak height. In other words the resolving power is equal to the FWHM from the peak width definition or the Δm from the valley definition.

The IUPAC description of the mass resolving power in mass spectrometry (M03730) refers directly to the valley (10% is recommended, and the used valley must always be stated) for two peaks (m1, m2) of equal height, as shown in equation 5:

$$m1 / (m1 - m2) \quad (5)$$

There is thus a deep inconsistency in these terms of resolving power. While the resolving power in R05321 is expressed as the peak width (at a certain fraction), the resolving power in M03730 is expressed as the ratio of the peak mass over Δm. But, if m=m1 and $m\text{-}\Delta m = m2$, the same ratio was already referred to as the resolution (in R05318), as shown in equation 6:

$$Resolution = m / \Delta m = m1 / (m1\text{-}m2) \quad (6)$$

So what does the resolving power really mean? Is it the peak width or the ratio? The literature does not provide a conclusive explanation. Boyd et al. describe the resolving power as a property of the instrument and the resolution as "*the separation between similar m/z values actually achieved in a real mass spectrum*" [4]. Resolving power is then explained using both the valley and the peak width definitions, where the resolving power is the ratio from equations 2 or 5. Even the IUPAC sponsored project to update the Standard Terms and Definitions for Mass Spectrometry [47] (so-called Mass Spec Terms Wiki), reports the term Resolution in mass spectrometry as a problematic term. The main part of the literature [21-24] in the mass spectrometry field uses resolution as the ratio $m/\Delta m$ and refers to the IUPAC definition R05318. However, in other fields of physics and chemistry, the ratio $m/\Delta m$ *is* usually described as the resolving power and Δm as the resolution, which has also been adopted in some mass spectrometry literature [25-27]. In the IUPAC recommendations this definition is consistent for microscopy and optical spectroscopy. Ken Busch [44] recommended to use the meaning common in most fields, where resolution is the difference Δm, while the resolving power it is the ratio R. The strong confusion between these two terms could be easily avoided by distinct and clear definitions.

4. Discriminability and distinguishability

Two other terms that are frequently encountered in MS, are the terms of discriminability and distinguishability. In order to investigate these terms, as well as to relate the terms to the ones already proposed, it is natural to include terms from the field of topology. Topology is one of the

unified branches of mathematics, and it is the study of qualitative characteristics of spaces. While topology generalizes shapes via abstraction, it also offers more formal definitions to describe some structural characteristics.

In topology, the separability is defined as the ability to divide the measurable space into countable dense subsets [47]. Every topological space is already dense in itself. Therefore, the 10% valley describes one of many possible designs of mass spectrum separation and is designated as resolution.

Additionally, topological discriminability is the quality to perceive or discern differences between two similar objects. While the mass spectral values represent the coordinates in two domains (mass and intensity), the shape of each mass-intensity pair could be considered as a point in topological space. Therefore, discriminability quantifies not only the position, but also the distance.The formal definition of discriminability is exhausting, but the interpretation is rather simple: "*Two objects are discriminable if there is an open sentence that is satisfied by one of the objects and not the other. If all the objects of domain are discriminable, then each of them uniquely satisfies infinite conjunction. Each real number is uniquely determined by the set of all the sentences that it satisfies. Ordinal numbers are only moderately discriminable, since any two of them satisfy the open sentence in one order an not the other*" [48, 49]. An open sentence is usually an equation or equality whose true value is meaningless until its variables are replaced with specific numbers. The mass values are ordinary numbers where the preceding relation is defined (m/z 100 < m/z 101). The preceding relation by itself already fulfills the existence of an open sentence. In other words, if it is possible to define some metric in which the two objects are discriminable (like relation/operator of < or >, or just ≠), then the objects are discriminable. It is not important *in what* they are discriminable, it is important that such metric could be defined.

In topology, distinguishability is usually considered to be the same concept as resolution, but mathematically it is a slightly different term: a set of non-empty values is required to distinguish between two values. In other words, two distinguished values have to share the same neighborhood value(s). There has to be a valley between the values. This definition of distinguishability represents a very important concept: it is describing the practically achieved ability, meaning that we can distinguish two mass peak (centroid) values only if there is a valley between them. Moreover, the IUPAC recommendation additionally puts requirements on the value of the valley. Thus, the IUPAC recommendation is a guideline on how to move from the theoretical description to practical values. The resolution is the theoretical characterization. According to Ken Busch's suggestion [44], resolution Δm is the theoretical limit of the distance between two points. Resolving power is the theoretical ability to resolve $R = m/\Delta m$ peaks of the 5% peak width equals Δm in the range $(0,m)$.

While distinguishability can be achieved in practice, the resolution is a theoretical potency estimated via calibrations. The practical impact is immediate. The theoretical resolution equals the distinguishability only in the ideal case. The distinguishability for real measurements is usually worse than the theoretical resolution.

It is a normal situation that the superimposed signal (as in figure 6i) does not show distinguishable peaks, even if the single peaks could be measured. This does not mean that the resolution gets worse, but the distinguishability of this particular case is worse than the resolution. Moreover, the superimposed signal of figure 6k cannot be clearly used for estimation of the resolution, but the two peaks are distinguishable - there is a valley. The concepts of discriminability and distinguishability create differences between mass values of the same peak and mass values of the other mass peaks.

Topological definitions discussed here are not in opposition to the previous descriptions. In contrast, they complete the resolution concept and repair some of the existing confusion. In real measurements, many of these valleys could be considered not valid: they could be caused by noise, or they could be not fulfilling some given criterion. This theoretical criterion is the resolution (10% valley, 5% peak width, or 2 FWHM). This leads us to the important difference between the theoretical and practical ability to distinguish adjacent MS peaks:

- All theoretical values are discriminable. The subsets of values (ideal peaks) fulfilling to be

isolated (5% criterion) or with a certain valley (10% criterion) are defining the value of the resolution. The resolution is a special case of distinguishability. The values with a distance between the points smaller than the resolution are just discriminable. The values with a bigger distance are distinguishable by the resolution. The detector has the resolving power to separate peaks (subset) of a given width (=resolution).

- All measured values are discriminable. The theoretical resolution define the *minimal* distance (discriminability value) when the points **may** become distinguishable. The criterion defined for resolution does not have to be fulfilled for the real peaks to be able to distinguish them (Figure 6k). It is enough that the distance between the apexes fulfill the value of theoretical resolution for isolated symmetrical peaks (theoretical), and the valley exist. The resolution is the minimal acceptable distinguishability, as shown in equation 7:

$$Discriminate \leq Resolution \leq Distinguishable. \quad (7)$$

Figures 6j and 6l show cases where peaks are not distinguishable, but where the distance between theoretical apexes equals the theoretical resolution.

4. Conclusions

Generally, the challenge of determinations of mass centroids consists of three related parts:

1) Defining the fundamental characteristics of MS
2) Establishing criteria on how to link the theoretical and practical ability of an instrument to distinguish adjacent MS peaks
3) Estimating the input parameters of the initial processing functions to obtain centroid positions and peak areas

A sufficient number of processing methods to decompose the mass spectrum into individual mass peaks and compute the centroids values are available. The crucial issue is, however, selecting input parameters for the initial processing functions. The scaling parameter or window length chosen is much more important for the result of the subsequent data analysis than the decision on the type of processing function.

The fundamental characteristics of resolution and resolving power are often described in different ways. The interpretation as well as the relations between different concepts is not always immediate. The IUPAC descriptions are only recommendation or statements instead of definitions. The resolution could be estimated on isolated peaks or valley fulfilling the 10% criterion. Otherwise, with two adjacent and overlapping peaks, it is recommended to estimate the FWHM. The FWHM has to be at least 1/2 of the theoretical resolution. In addition some valley has to be present between two peaks. The IUPAC 10%, 5%, and 50% fraction definitions, respectively, constitute recommendations on how to estimate the scale parameters. The relations between valley Δm, $5\% \Delta m$, FWHM scale parameter are known.

The terms of distinguishability and discriminability complete our recommendations of fundamental definitions in MS. Topological definitions offer a decision between the theoretical ability of a measurement device and the practically achieved ability of the measurement to distinguish adjacent MS peaks. If there is a valley between the mass peaks, and the distance between mass peak apexes is bigger than theoretical resolution, then the peaks are distinguishable. In other words, the distinguishability cannot be better than the resolution.

Resolution and resolving power are theoretical values for the ideal case. In real measurements we have to deal with the distinguishability. The minimal possible value of the distinguishability is given by the theoretical resolution. The value of distinguishability could be estimated by the use of 2 FWHM.

Acknowledgment

Authors are grateful to Anastasia Hole, Ibrahim Karaman, Guro Dørum, Gesine Schmidt, Jan Vaněk, Pavla Urbanová, and Rebecca Smaha for relevant discussion. This work was supported by the Research Council of Norway, under the Yggdrasil project 210955/F11; by CENAKVA CZ.1.05/2.1.00/01.0024; by the South Bohemia University grants GAJU 152/2010/Z and GAJU 134/2013/Z; and by the project Postdok JU CZ.1.07/2.3.00/30.0006. Financial support from the Agricultural Food Research Foundation of Norway is also greatly acknowledged.

References

1. V. Lindberg. Uncertainties, Graphing, and the Vernier Caliper. Rochester Institute of Technology, (2003).
2. K.H. Liland; Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis, Trends in Analytical Chemistry, Vol. 30, No. 6, 2011.
3. A.G. Stephenson; L.S. LaPiana.; D.R. Mulville; P.J. Rutledge; F.H. Bauer; D. Folta; G.A. Dukeman; R. Sackheim; et al. Mars Climate Orbiter Mishap Investigation Board Phase I Report. NASA (1999).
4. R.K. Boyd, C. Basic, R.A. Bethem. Trace Quantitative Analysis by Mass Spectrometry. Wiley (2008); p: 8, 249, 256.
5. J. Cox; M. Mann. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. Journal of The American Society for Mass Spectrometry 20(8). (2009). 1477-1485.
6. M.C. McMaster: HPLC, a practical user's guide; Wiley, 2007.
7. R.E. Ardrey. Liquid Chromatography Mass Spectrometry: An Introduction; Wiley, 2003.
8. D. Noble. The Music of Life: Biology beyond genes; Oxford University Press, 2006.
9. M. Katajama, Orešič, M., Data processing for mass spectrometry-based metabolomics, Journal of Chromatografhy A, 1158, p. 318-328, 2007.
10. J. Urban. J. Vaněk. J. Soukup. D. Štys, Expertomica metabolite profiling: getting more information from LC-MS using the stochastic systems approach. Bioinformatics. Oct 15;25(20):2764-7.Vol. 25 no. 20 2009, pages 2764–2767 (2009).
11. Compendium of Analytical Nomenclature (definitive rules 1997), 3rd edition Inczedy, J.; Lengyel, T. and Ure, A.M. Blackwell Science, 1998 [ISBN 0-86542-6155] , on-line version: http://old.iupac.org/publications/analytical_compendium Web site constructed by David S. Moore, Last updated 31 July 2002.
12. E.R. Cohen, T. Cvitas, J.G. Frey, B. Holmström, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H.L. Strauss, M. Takami, and A.J. Thor, "Quantities, Units and Symbols in Physical Chemistry", IUPAC Green Book, 3rd Edition, 2nd Printing, IUPAC & RSC Publishing, Cambridge (2008)
13. R.G. Cooks; A. L. Rockwood. 'The 'Thomson'. A suggested unit for mass spectroscopists. Rapid Communications in Mass Spectrometry 5. (2): 93. (1991).
14. Bureau International des Poids et Mesures. Le système International d'Unités (SI). 8th French and English Edition, BIPM, Sèvres, 2006.
15. R.C. Murphy; S. J. Gaskell; New Applications of Mass Spectrometry in Lipid Analysis; J. Biol. Chem. (paper in press), doi/10.1074/jbc.R111.233478. 2011.
16. Author Guidelines, Rapid Communications in Mass Spectrometry. Wiley Interscience. 2012.
17. T. H. Bertram, J. R. Kimmel, T. A. Crisp, O. S. Ryder, R. L. N. Yatavelli, J. A. Thornton, M. J. Cubison, M. Gonin, and D. R. Worsnop; A field-deployable, chemical ionization time-of-flight mass spectrometer: application to the measurement of gas-phase organic and inorganic acids; Atmos. Meas. Tech. Discuss., 4, 1963-1987, 2011.
18. W.X. Schulze; B. Usadel; Quantitation in Mass-Spectrometry-Based Proteomics; Annu. Rev. Plant Biol. 61:491–516. 2010.
19. N. E. Manicke, A. L. Dill, D. R. Ifa, R. G. Cooks; High-resolution tissue imaging on an orbitrap mass spectrometer by desorption electrospray ionization mass spectrometry; Journal

of Mass Spectrometry; Volume 45, Issue 2, pages 223–226, February 2010.
20. B.N. Pramanik, M.S. Lee, G. Chen, S.A. Smith, R.W. Smith,Y. Xia, Z. Ouyang; Chapter Introduction to Mass Spectrometry in Characterization of Impurities and Degradants Using Mass Spectrometry; Willey, 2011.
21. F.W. McLafferty, F. Tureček; Interpretation of Mass Spectra, University Science Books, 1993.
22. E. de Hoffmann, V. Stroobant, Mass Spectrometry: Principles and Applications; Wiley-Interscience, 2007.
23. S. Sechi, Quantitative Proteomics by Mass Spectrometry (Methods in Molecular Biology), Humana Press, 2007.
24. I. Eidhammer, K.Flikka, L. Martens, S.-O. Mikalsen, Computational Methods for Mass Spectrometry Proteomics, Wiley-Interscience, 2008.
25. O.D. Sparkman, Mass Spectrometry Desk Reference, Global View, 2006.
26. J.T. Watson, O.D. Sparkman, Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation; Wiley, 2007.
27. Ch. Dass, Fundamentals of Contemporary Mass, Wiley - Interscience, 2007.
28. A.J. Den Dekker, A. van den Bos; Resolution: a survey; J. Opt. Soc. Am. A, Vol. 14, No.3, 1997.
29. J.Urban, An Induction to the Theory of Stochastic Systems, Lectures, http://www.auc.cz/ipb/vpk/doc/stochastika112010/stochastika_urban_prez.pdf, University of South Bohemia, 2010.
30. T. Mallet, Elemental composition from accurate m/z determinations, Essay, University of Greenwich, 2005.
31. A.H. Grange, W.C. Brumley, A Mass Peak Profile Generation Model to Facilitate Determination of Elemental Compositions Based on Exact Masses and Isotopic Abundances, Journal of the American Society for Mass Spectrometry, Volume 8, Issue 2, Pages 170–182, 1997.
32. A.G. Marshall, C.L. Hendrickson, G.S. Jackson, Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer, Mass Spectrometry Reviews, 17, 1–35, 1998.
33. Polettini, A. (ed.), Applications of LC-MS in Toxicology; Pharmaceutical Press, 2006.
34. Ledvij, M, Curve Fitting Made Easy; The Industrial Physicist pp. 24- 27,2003.
35. Reed, J., Curve Fitting; Lessons on Introduction to Statistics and Probability, http://argyll.epsb.ca/jreed/, 2000.
36. IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook. Last update: 2011-10-11; version: 2.3.
37. Chi Woo, Aaron Krowne. "mean square error" (version 7). PlanetMath.org. Freely available at http://planetmath.org/?op=getobj;from=objects;id=1289
38. Savitzky, A.; Golay, M.J.E. (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". Analytical Chemistry 36 (8): 1627–1639. doi:10.1021/ac60214a047
39. P.O. Persson, G. Strang, Smoothing by Savitzky–Golay and legendre filters, Comm. Comp. Finance, 13 (2003), pp. 301–316
40. M. Browne, N. Mayer, T.R.H. Cutmore, A multiscale polynomial filter for adaptive smoothing, Digital Signal Processing, Volume 17, Issue 1, January 2007, Pages 69–75.
41. Mass Spectrometry Terms and Definitions Project Page, http://mass-spec.lsu.edu/msterms, last modified on 23 August 2011.
42. John Brignell, Number Watch, Brignell Associates, Mere, Warminster, 2006.
43. Christie G. Enke, The science of chemical analysis and the technique of massspectrometry, International Journal of Mass Spectrometry, (2011), V212, I1–3, pp1-11.

44. K. L. Busch, T.A. Lehman, Guide to Mass Spectrometry, Vch Pub, 2000.
45. The HPLC Troubleshooting Wizard, LC Resources , 2006.
46. http://mass-spec.lsu.edu/msterms/
47. J.C.R. Alcantud, Topological separability and axioms of countability in GPO-spaces, Bull. Austal. Math. Soc., (1997), V. 55, pp. 131-142.
48. W. V. Quine, Grades of Discriminability, The Journal of Philosophy, (1976) Vol. 73, No. 5, pp. 113-116.
49. Eric Hoekstra, Henk Wolf, The Principle of Distinctivity, Linguist List (2005), 16.1608.
50. Sturm et al., BMC Bioinformatics (2008), 9, 163., Kohlbacher et al., Bioinformatics (2007), 23:e191-e197.
51. T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics 11:395 (2010).
52. N.Markevich, I.Gertner, Comparison among methods for calculating FWHM, Nuclear Instruments and Methods in Physics Research, (1989), A283, pp 72-77.

Figure 1. A typical 3D representation of a blank LC-MS measurement using TopView 1.8 software [50] (left plot). The Total Ion Current chromatogram (TIC, upper right plot) and a selected mass spectrum (lower right plot) are obtained using Mzmine 2.2. [51]

Figure 2. Illustration of differences between profile (top) and centroid (bottom) data obtain with an ion trap detector (pig blood), a quadrupole detector (beer), a QTOF detector (phenolic acids), and an Orbitrap detector (MeOH), respectively.

Figure 3. Peak shape fitting of two mass peaks, coumaric acid (left) and Sinapic acid (right), using Gaussian and Laplacian distribution fitting, respectively. Data were obtained by LC-MS with QTOF detection.

Figure 4. Mass spectrum smoothing by window functions.
A: Raw mass spectrum (beer) smoothed using a Gaussian apodization filter (scale parameter : 0.4 *m/z* units)
B: Comparison of Gaussian and triangularfitting using the same scale parameter (0.4 *m/z* units).
C: Comparison ofGaussian fitting with different window lengths.

Figure 5. Estimation of mass peak centroid positions using wavelet transformation (right) of raw signals (left).

Figure 6. Illustrations of the IUPAC R05318 recommendation and the relationship between the peak valley and peak width definitions. The left column explains the analogy of 5% fraction and 10% peak valley for peaks of equal heights. The middle column extends the analogy also for peaks of non-equal heights. The right column explains the practical difference between the resolution evaluated via peak valley and full width at half maximum (FWHM). Additional details are described in the main text.

R05318	resolution in mass spectroscopy:
energy	By analogy with the peak width definition for mass resolution, a peak showing the number of ions as a function of their translational energy should be used to give a value for the energy resolution.
10 per cent valley definition	Let two peaks of equal height in a mass spectrum at masses m and $m - \Delta m$ be separated by a valley which at its lowest point is just 10 per cent of the height of either peak. For similar peaks at a mass exceeding m, let the height of the valley at its lowest point be more (by any amount) than ten per cent of either peak height. Then the resolution (10 per cent valley definition) is $m / \Delta m$. It is usually a function of m. The ratio $m / \Delta m$ should be given for a number of values of m.
peak width definition	For a single peak made up of singly charged ions at mass m in a mass spectrum, the resolution may be expressed as $m / \Delta m$ where Δm is the width of the peak at a height which is a specified fraction of the maximum peak height. It is recommended that one of three values 50%, 5% or 0.5% should always be used. For an isolated symmetrical peak recorded with a system which is linear in the range between 5% and 10% levels of the peak, the 5% peak width definition is technically equivalent to the 10% valley definition. A common standard is the definition of resolution based upon Δm being Full Width of the peak at Half its Maximum height, sometimes abbreviated 'FWHM'. This acronym should preferably be defined the first time it is used.

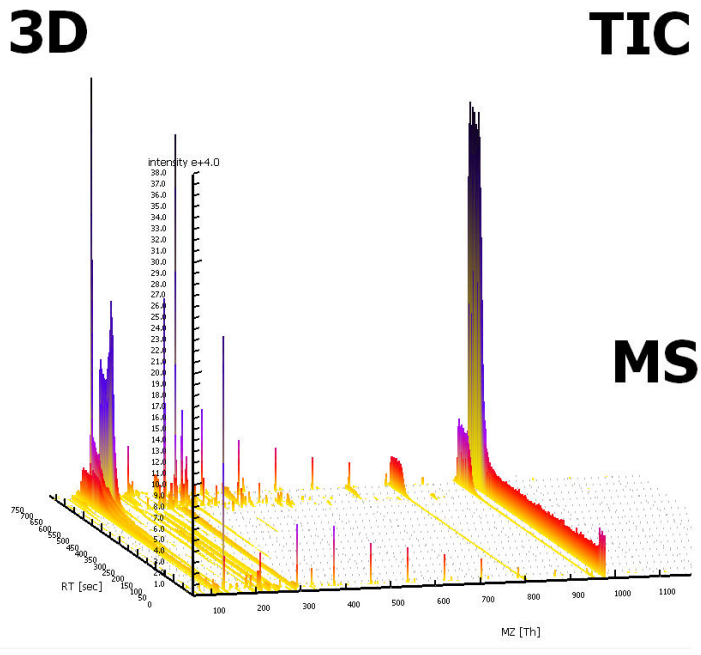
Table 1: Three IUPAC recommendations for the term resolution in mass spectrometry [50]:the so-called energy definition, the peak valley definition, and the peak width definition.

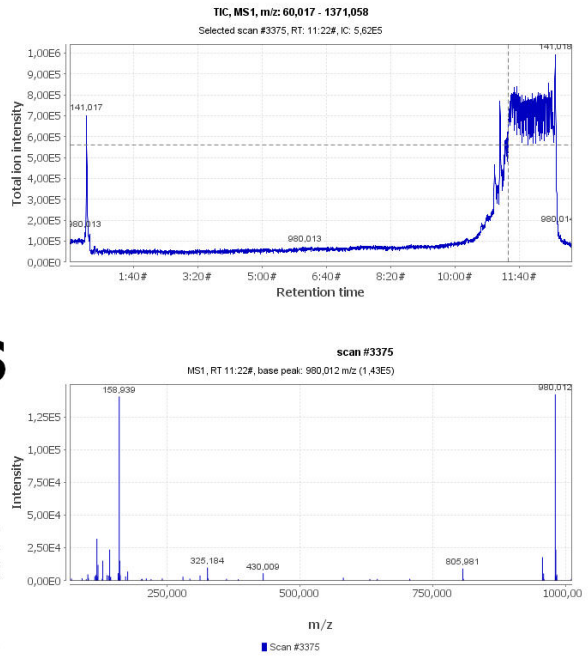
R05321	resolving power in mass spectrometry
	The ability to distinguish between ions differing in the quotient mass/ charge by a small increment. It may be characterized by giving the peak width, measured in mass units, expressed as a function of mass, for at least two points on the peak, specifically at fifty percent and at five percent of the maximum peak height.
M03730	mass resolving power in mass spectrometry
	Commonly and also acceptably defined in terms of the overlap (or 'valley') between two peaks. Thus for two peaks of equal height, masses m_1 and m_2, when there is overlap between the two peaks to a stated percentage of either peak height (10% is recommended), then the resolving power is defined as $m_1/(m_1 - m_2)$. The percentage overlap (or 'valley') concerned must always be stated.

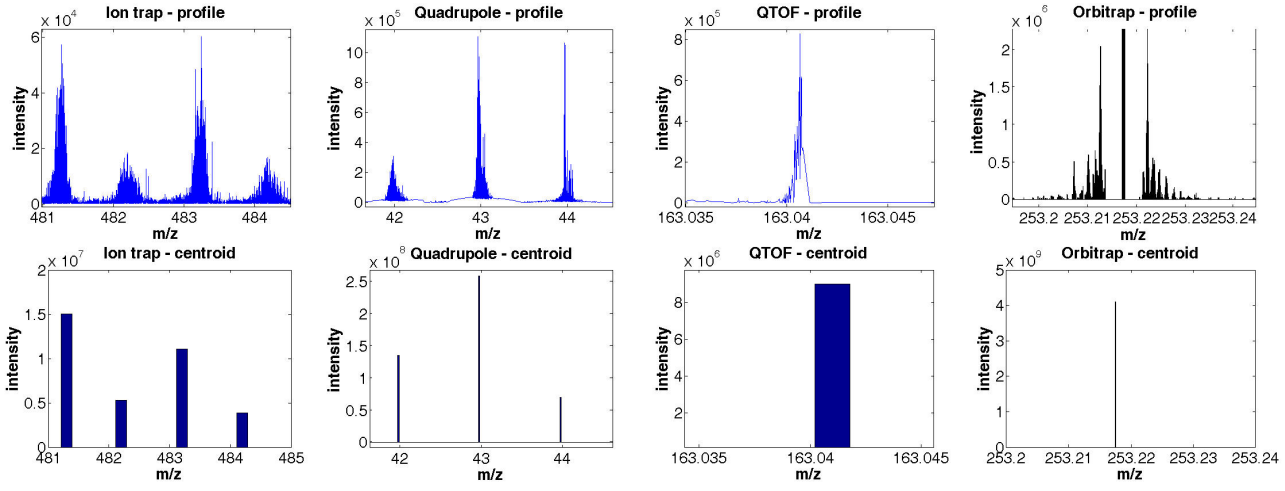
Table 2: IUPAC recommendations for the terms resolving power in mass spectrometry (R05321) and mass resolving power in mass spectrometry (M03730) [50].

Peak shape	Multiplicator
Blackman	0,810957
Lorentzian	2
Gaussian	2 sqrt(2 ln(2))
Hamming	1,05543
Bartlett	1

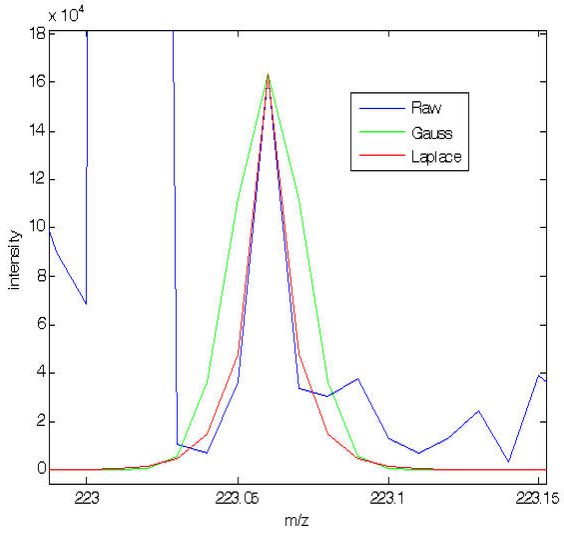
Table 3: Relation of full width at half maximum (FWHM) and scale parameter of five common distribution functions (http://mathworld.wolfram.com/FullWidthatHalfMaximum.html).

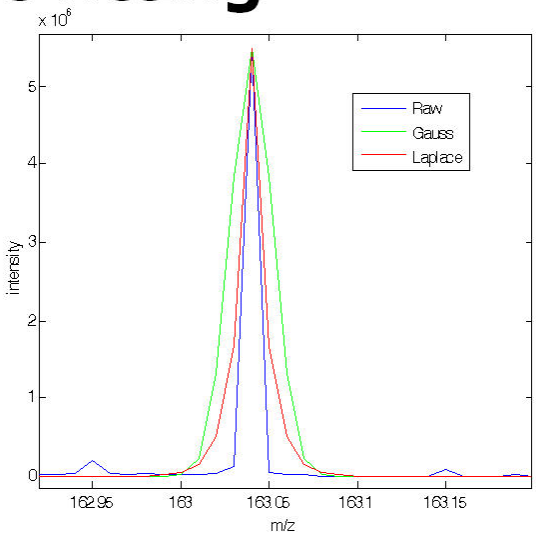

3D
intensity e+4.0
RT [sec]
MZ [Th]

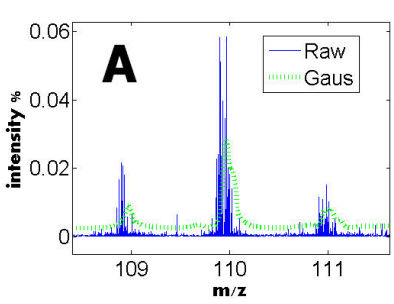

TIC
TIC, MS1, m/z: 60,017 - 1371,058
Selected scan #3375, RT: 11:22#, IC: 5,62E5
Total ion intensity
Retention time
MS
scan #3375
MS1, RT 11:22#, base peak: 980,012 m/z (1,43E5)
Intensity
m/z
Scan #3375

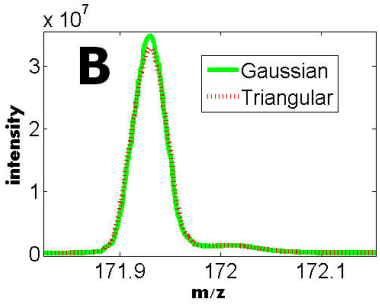

Ion trap - profile
Quadrupole - profile
QTOF - profile
Orbitrap - profile
Ion trap - centroid
Quadrupole - centroid
QTOF - centroid
Orbitrap - centroid
intensity
m/z

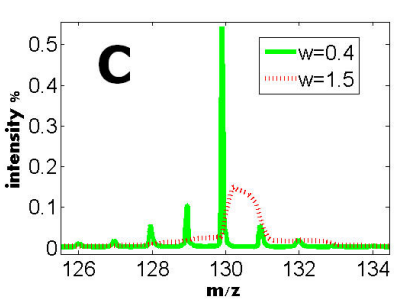
Peak shape fitting











Centroiding via wavelet transformation