Published in final edited form as:
J Am Soc Mass Spectrom. 2011 November ; 22(11): 2090–2099. doi:10.1007/s13361-011-0229-2.

A practical approach for determination of mass spectral baselines

Kui Yang[1,*], **Xiaoling Fang**[2,*], **Richard W. Gross**[1], and **Xianlin Han**[2,#]

[1]Division of Bioorganic Chemistry and Molecular Pharmacology, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110

[2]Sanford-Burnham Medical Research Institute, Orlando, FL 32827

Abstract

Precise determination of the baseline levels of mass spectra is critical for identification and quantification of analytes. Herein, we present a practical approach for determination of the baselines of mass spectra acquired under differential conditions. The baseline determined by this approach was the sum of baseline drift and noise level. The baseline drift was determined by averaging a number of lowest ion intensities. The noise level was determined based on the fact that an accelerated intensity change exists from noise to signal. This change was best revealed by the established accumulative layer thickness curve that was derived from the thicknesses of individual deducted layers. Deductions were performed sequentially layer by layer each of which has a thickness of averaged lowest ion intensities from existing spectral data. The layer where the accelerated intensity change occurred was defined as a transition layer which was determined from the polynomial regression in the sixth order of the accumulative layer thickness curve followed by resolving the roots of its fourth derivative. We validated the presence of this transition layer through determination of its convergence from various accumulative layer thickness curves generated by varying either the ending or the fineness of the sequential layer deductions. This simple, practical, program-based baseline determination approach should greatly increase the accuracy and consistency of identification and quantification by mass spectrometry, and facilitate the automation of data processing, thereby increasing the power of any high throughput methodology in general and of shotgun lipidomics in particular.

Keywords

baseline correction; mass spectrometry; shotgun lipidomics

Introduction

Determination of a baseline is apparently critical for identification of analytes by mass spectrometry through properly discriminating noises and signals. It is also essential for accurate quantification of the mass content of each analyte through correcting the baseline contribution to the individual peak intensities of a mass spectrum, particularly when a shotgun lipidomics approach is employed where quantification of analytes is performed through direct comparison of their ion intensities with that of selected internal standard(s)

[#]To whom correspondence should be addressed: Xianlin Han Sanford-Burnham Medical Research Institute 6400 Sanger Road Orlando, FL 32827 Tel.: 407-745-2139 Fax: 407-745-2013 xhan@sanfordburnham.org.

[*]These authors equally contribute to the study.

Appendix A

Supplementary Materials The supplementary materials associated with this article are attached.

[1] and when the species is in low abundance (e.g., signal/noise < 10). Baseline correction is also important for precise display of mass spectra, for example, in tissue mapping [2].

Two major factors contribute to the values of a baseline of a mass spectrum, i.e., detector drift (i.e., baseline drift) and chemical noise (i.e., noise level). The baseline drift represents a shift of a mass spectrum from the origin (i.e., zero) and can be tuned to a linear, minimal increase with increase in *m/z* for a mass spectrometer with a quadrupole as an analyzer. The chemical noise, the actual noise level of a spectrum after deduction of the baseline drift, is composed of background signals resulting from residual chemicals. In particular, inorganic salts presented in the sample matrix are largely responsible for this noise.

Many different approaches, based on a variety of algorithms, have previously been developed and been applied for different purposes [2-4]. However, most of these approaches have been limited to their particular utilities (e.g., correction for matrix-induced baseline in matrix-assisted laser desorption/ionization mass spectrometry; correction for mobile phase components-contributed baseline; etc.). Herein, we present a simple, practical approach for a general determination of the baseline of a mass spectrum. This approach includes the corrections for both the baseline drift and the noise level of a mass spectrum. The principles of this practical approach are discussed and the procedures to determine the baseline are described in detail. We compared many of the baselines determined using either this program-based approach or a manual approach and found that the baselines determined by these two approaches were very consistent. We specifically pointed out that although the new approach was derived largely from the spectra acquired by the Xcalibur operation system, it could be readily modified for the application for other operation systems as we practiced for the mass spectra acquired by ABI 4800 mass analyzer. We believe that the development of this simple, practical, program-based approach should greatly facilitate the automation and increase the power of high throughput methodology.

Experimental

Determination of the Baseline Drift of a Mass Spectrum

In theory, the baseline drift of a single spectral peak acquired in the profile mode can be corrected by deduction of the lowest intensity of its data points. The baseline drift of a mass spectrum in a certain mass range can thus be corrected by averaging the lowest intensities of individual peaks of the mass spectrum (Scheme 1A). In practice, we first determined the number of peaks (N) of the mass spectrum and then averaged the N lowest intensities from the spectral data. This averaged intensity was defined as the baseline drift of the spectrum and deducted from the spectral data (Scheme 1B). The detailed procedures were described below.

Each of the mass spectra processed in the current work were acquired in the profile mode by the Xcalibur operation system. By using the Xcalibur Qual Browser, each spectrum was averaged from the acquired individual scans and smoothed with a Gaussian function. A full list of the spectral data points within a user-specified mass range was generated in the Qual Browser with the display option specified for “All peaks”. In this data list, the *m/z* difference between every two neighboring data points was one tenth of the Peak Width (FWHM). For example, when a unit resolution spectrum was acquired at a Peak Width setting of 0.7 Th [5, 6], the *m/z* difference between two neighboring data points was 0.07 Th. Each single peak in a unit resolution spectrum was then composed of 14-15 data points (i.e., 1/0.07 = 14.3). If a spectrum was acquired at a Peak Width setting of 0.35 Th, the *m/z* difference between two neighboring data points was 0.035 Th and the spectrum was a half unit resolution spectrum each peak of which occupied a half mass unit and was also composed of 14-15 data points (i.e., 0.5/0.035 = 14.3). Accordingly, we determined the number of peaks of a mass spectrum from the number of raw spectral data points by the following equation:

$$\begin{aligned}\text{Number of peaks}\quad(\text{N})\quad\text{of a spectrum}\quad&=\text{The total number of raw data points of the spectrum}/14.3\\&=\text{The total number of raw data points of the spectrum}*0.07\end{aligned}\qquad(1)$$

For example, a unit resolution spectrum of a full list of 1430 data points contained 100 peaks (i.e., N = 1430 * 0.07 = 100). If, within the same mass range, a half unit resolution spectrum was acquired, a full list of 2860 spectral data points would be generated because the mass difference (i.e., 0.035) between every two neighboring data points was half of that from unit resolution spectrum. The number of peaks of the spectrum would therefore be doubled (i.e., N = 2860 * 0.07 = 200).

Next, we sorted the data in an ascending order of the intensities, and then averaged the N lowest intensity data. This averaged intensity was defined as the baseline drift of the mass spectrum in the mass region. For example, the baseline drift of the unit mass spectrum exemplified above was obtained by averaging the 100 lowest intensities after sorting the intensity data while the baseline drift of the exemplified half unit mass spectrum was obtained by averaging the 200 lowest intensities. Finally, the baseline drift was subtracted from each of the raw spectral intensity data points. Those data points that had intensities of either zero or negative after the baseline drift deduction were discarded. The remaining data points were re-sorted in an *m/z* order and termed as **Data Set 0** which represented the baseline drift-deducted data set. A mass spectrum reconstructed from the baseline drift-deducted data set (i.e., Data Set 0) had baseline drift deducted but still had background noise present (Figures S1 to S4).

Determination of the Noise Level of a Mass Spectrum

The baseline drift-deducted data points (i.e., Data Set 0) of a mass spectrum were next used for determination of the noise level of the spectrum. In theory, the same procedure as the baseline drift deduction described above could be repeated on Data Set 0, which in turn generates a new data set (i.e., **Data Set 1**). If this type of deduction procedure is repeated over and over, many Data Sets could be generated from their previous ones (e.g., Data Set M from Data Set M-1) where each Data Set has less data points compared to its previous one. The mass spectrum reconstructed from Data Set 1 could be imagined as a spectrum with a thin “**layer**” wiped off from the bottom of the baseline drift-deducted spectrum (reconstructed from Data Set 0) while **the thickness of the “layer”** is the average of N (calculated by Equation 1) lowest intensities from Data Set 0. The mass spectrum reconstructed by Data Set M could be imagined as a spectrum with a “layer” wiped off from the bottom of the spectrum reconstructed from Data Set M-1 while the thickness of the “layer” is the average of N lowest intensities from Data Set M-1.

The thickness of each “layer” may vary. A “layer” wiped off to generate a very early Data Set (e.g., Data Set 1) from its previous Data Set (i.e., Data Set 0) is usually thin, while the “layer” wiped off to generate a later Data Set from its previous Data Set could be much thicker. One of the reasons is that each peak contains the same number of data points (i.e., 14-15) equally distributed in *m/z* dimension (i.e., x-axis), thus the higher the intensity of the peak, the bigger the intensity difference between its two neighboring data points (Scheme 1A). Since the spectral peaks are Gaussian smoothed and of Gaussian distribution, this intensity difference increases with the data point moving towards the top of the peak in comparison to the difference between two points near the bottom of the peak (Scheme 1A). Another reason is that the same number (N, calculated by Equation 1) of lowest intensities is averaged from each Data Set to generate next Data Set. Accordingly, the very early layer by layer deductions would not lead to a significantly-varied thickness of each layer since the N lowest intensities in each Data Set are almost exclusively from the low intensity peaks (i.e., noise peaks) and the intensity differences between neighboring data points of the low intensity peaks are small. When low intense peaks are wiped off after a few times of deduction, the data points from high intensity peaks (or signals) are then picked up in the N lowest intensities whose average is the layer thickness for deduction from the current Data Set to generate next Data Set. Therefore, a significant increase in layer thickness would occur at this point. We defined this layer as the **Transition Layer**. The sum of the deducted intensities (or the layer thicknesses) of individual layers from the first layer to the transition layer was designated as **the noise level of the spectrum**, which was deducted from Data Set 0 to obtain the spectrum with both baseline drift and noise corrected.

In reality, when working on complex mass spectra, searching for the transition layer might not be as straightforward as described above. We performed the following steps to practically search for this transition layer from the baseline drift-deducted spectral data points (i.e., Data Set 0).

1. Calculation of the thickness of a layer

To calculate the thickness of each layer, we repeated the steps for determining the baseline drift described above on every new Data Set except that the number (N) of lowest intensity data points was replaced with a new number (n) which was calculated by the following equation:

$$n=\text{The total number of raw data points of a spectrum }^{*}\text{ Step Length} \tag{2}$$

where the **Step Length** determined how finely each layer deduction would be processed and is defined by the user. This number (n) could be different from the number (N). If a step length of 0.07 was defined, the number (n) calculated by Equation 2 was identical to the number (N) by Equation 1. If a step length of < 0.07 was defined, a smaller number of lowest intensity data points from the current Data Set were used for deduction to yield the next Data Set. The average of a smaller number (e.g., p) of lowest intensity data points (e.g., intensity 1, intensity 2, ..., intensity p in an ascending order) was smaller than the average of a bigger number (e.g., q, q > p) of lowest intensity data points (e.g., intensity 1, intensity 2, ..., intensity p, intensity p+1, ..., intensity q in an ascending order). Accordingly, this represented a finer sequential layer deduction because more layers of deduction would necessarily be performed to wipe off the entire spectral data points if a smaller averaged intensity was used for deduction of each layer. In contrast, the defined step length of > 0.07 led to a rougher layer deduction represented by a bigger averaged intensity for deduction of each layer and less layers of deduction for the entire spectrum. The significance of varying the number (n) for deduction of each layer was discussed in detail in the Discussion section.

In practice, we first sorted the data points from the baseline drift deducted spectral data (Data Set 0) in the order of ascending intensity and averaged the n lowest intensity data points (n was calculated by Equation 2 with a user-defined Step Length). Then we deducted this averaged intensity from the intensities of individual data points in Data Set 0 and discarded the data points whose intensities were zero or negative after deduction. The remaining data points yielded Data Set 1. The mass spectrum reconstructed from the newly generated Data Set 1 could be viewed as the baseline drift-deducted spectrum (reconstructed from Data Set 0) was wiped a layer off from the bottom of the spectrum whose thickness was the average of the n lowest intensities from Data Set 0. This averaged intensity was designated as the **_T_hickness of _L_ayer** 1 (TL_1). This procedure was repeated to calculate the thicknesses of sequential layers. In general, the calculation of the Thickness of Layer i (TL_i) could be represented by the following equation:

$$\mathrm{TL_i{=}Average\,(Intensity\quad 1, Intensity\quad 2, \ldots, Intensity\quad n\quad of\ Data\ Set\quad i{-}1)} \tag{3}$$

where TL_i is the Thickness of Layer i, and Intensity 1, Intensity 2, ..., Intensity n are the n lowest intensities of Data Set i-1. The deduction of TL_i from the data points of Data Set i-1 yielded Data Set i. For example, Data Set 2 was generated from Data Set 1 by wiping from Data Set 1 one layer off with a thickness of TL_2 which was calculated from averaging the n lowest intensities of Data Set 1. The procedure was repeated until the number of the remaining spectral data points was smaller than n. A series of the thicknesses of layers: TL_1, TL_2, ..., TL_m were generated accordingly where TL_m was the thickness of the last Layer m, whose deduction generated the last Data Set m from Data Set m-1 while one-step further deduction on Data Set m would have generated a Data Set that contained less than n remaining spectral data points.

2. Generation of the accumulative thickness of a layer

In reality, it was difficult to directly determine the transition layer from the generated series of the layer thicknesses (i.e., TL_1, TL_2, ..., TL_m for layer 1, layer 2, ..., and last layer m, respectively). One of the reasons was that the curve of layer thickness (TL) *vs*. layer mostly had irregular trends for which no regression method worked consistently for different mass spectra with satisfactory correlation coefficients. To resolve this issue, we derived the ***A*ccumulative *T*hickness of a *L*ayer** (ATL) from the individual Thickness of the Layer (TL) by the following equation:

$$\mathrm{ATL_i{=}\sum TL_i} \tag{4}$$

where TL_i was the Thickness of Layer i (1 ≤ i ≤ m) and was calculated by Equation 3, and ATL_i was the Accumulative Thickness of Layer i. The curve of the Accumulative Thickness of Layer (ATL) *vs*. layer was termed as the **accumulative layer thickness curve**, which was used for determining the transition layer in the following steps.

3. Automated determination of the transition layer

Next, we fitted the accumulative layer thickness curve by regression to determine the transition layer. The first point (1, ATL_1) of the curve was not included for the curve fitting because the thickness of the very first layer (i.e, TL_1 or ATL_1) was lack of stability and the first layer impossible to be the transition layer. We determined the transition layer from the accumulative layer thickness curve by the following procedures:

i. Fitted these accumulative layer thickness data points by polynomial regression in the order of 6 by using MATLAB function polyval as follows:

$$y{=}a{+}bx{+}cx^2{+}dx^3{+}ex^4{+}fx^5{+}gx^6 \tag{5}$$

where "x" was the layer number (1 ≤ x ≤ m-1); "y" was the accumulative layer thickness (ATL) of the layer x, which is calculated from Equation 4; a, b, ..., and g were the regression coefficients. To specify, the first data point (x_1, y_1) for the regression is the second point (2, ATL_2) of the original accumulative layer thickness curve due to elimination of the first layer for regression, and the last data point (x_{m-1}, y_{m-1}) is the last point (m, ATL_m) of the original curve.

ii. Calculated the derivatives of the obtained polynomial regression up to the fourth derivative. The fourth derivative was as follows:

$$y'''' = 24e + 120fx + 360gx^2 \tag{6}$$

where e, f and g were regression coefficients from Equation 5.

iii. Found zeros of y’’’’ by solving the following single-variable quadratic equation by using MATLAB function polyder:

$$y'''' = 0 \quad \text{or} \quad 15gx^2 + 5fx + e = 0 \tag{7}$$

where e, f and g were regression coefficients from Equation 5. There were two possibilities for the roots of Equation 7: the two roots were real numbers x_I and x_{II} when $5f^2 - 12eg \geq 0$ or the two roots were complex numbers when $5f^2 - 12eg < 0$. When two real-number roots were obtained and $x_I \neq x_{II}$, the bigger number was taken as the transition layer and the smaller number was discarded. When two complex-number roots were obtained, the algorithm redid steps (i), (ii) and (iii) with the narrowed accumulative layer thickness curves by eliminating the last data point until real number roots were found from Equation 7.

4. Determination of the accumulative layer thickness corresponding to the determined transition layer

After the transition layer (e.g., the x_I from step 3) was determined, we calculated the accumulative layer thickness $\mathbf{y_I}$ corresponding to x_I. If x_I was an integer number, the y_I from the data point (x_I, y_I) of the accumulative layer thickness curve was then taken as the accumulative layer thickness corresponding to the transition layer x_I. If x_I was not an integer number, the “y”s from the two adjacent data points that were neighbors to the determined transition layer x_I (i.e., data points (x_s, y_s) and (x_{s+1}, y_{s+1}) where $x_s < x_I < x_{s+1}$) were used to calculate the y_I corresponding to x_I by the following equation:

$$y_I = y_s{}^* \; (x_{s+1} - x_I) + y_{s+1}{}^* \; (x_I - x_s) \tag{8}$$

where x_I is the determined transition layer; y_I is the accumulative layer thickness corresponding to x_I; x_s and x_{s+1} are the two adjacent layers from the accumulative layer thickness curve that meet $x_s < x_I < x_{s+1}$; y_s and y_{s+1} are accumulative layer thicknesses corresponding to x_s and x_{s+1}, respectively.

5. Self-check and determination of the spectral noise level

In theory, the determined accumulative layer thickness y_I corresponding to the determined transition layer x_I could be considered as the noise level of the spectrum. In practice, to self check the stability of the determined transition layer and eliminate any potential uncertainty from single time polynomial regression on selected data points, we determined more transition layers from narrower regions of the accumulative layer thickness curve. Specifically, we first repeated the procedures (i) (ii) and (iii) of Step 3 to individually determine the transition layers (e.g., x_I^1, x_I^2, ...) from a series of narrowed regions of the accumulative layer thickness curve (i.e., x = 1 to m-2, x = 1 to m-3, ..., and x = 1 to 7) which were yielded by eliminating the last data point of the curve sequentially until minimal number (i.e., seven) of data required for determination of the regression coefficients by Equation 5 reached. Then, the accumulative layer thicknesses y_I^1, y_I^2, ..., corresponding to the newly determined transition layers x_I^1, x_I^2, ..., respectively, were determined by repeating Step 4. We discarded the values from the determined y_I^1, y_I^2, ... that were either larger than y_I or less than y_I * 70% and averaged the rest of the values. This averaged accumulative layer thickness was defined as the noise level of a mass spectrum. The overall baseline of a mass spectrum was corrected by the sum of the determined baseline drift and the determined noise level of the spectrum. Meanwhile, the signal-to-noise ratio of an ion peak can be calculated as:

$$\begin{aligned}\mathrm{S/N}\ &=\text{the signal of an ion peak/the noise level of the spectrum}\\&=\text{baseline}-\text{corrected peak intensity/the noise level of the spectrum}\\&=(\text{ion peak intensity}-\text{the overall baseline})\ /\text{the noise level of the spectrum}\end{aligned}\tag{9}$$

Results and Discussion

Correction for Baseline Drift of a Mass Spectrum

The baseline drift of a mass spectrum is a constant shift of the peak intensities from their original values to the apparently determined values and consistently occurs for the entire spectrum. This drift is different from the noise background of the spectrum. The baseline drift of a mass spectrum could be either mass independent or mass dependent. However, when the mass range of interest is narrow, the baseline drift usually becomes minimally mass dependent within the mass range. In the current study, we focused on the correction of the mass independent baseline drift of a mass spectrum. When the mass range of interest is wide and the mass dependence of baseline drift has to be addressed, one can segment the entire widely ranged mass spectrum into a few narrowly ranged spectra [7] and then employ the current approach to correct the baseline drift individually for each of the segmented mass spectra.

The baseline drift of a Gaussian-smoothed single spectral peak acquired in the profile mode can be represented by its lowest intensity which is usually the intensity from either the first data point or the last data point of the peak. The baseline drift of a mass spectrum acquired in the profile mode containing N Gaussian-smoothed single peaks, in theory, can be represented by the line connecting each lowest intensity data point of each of the N single peaks in the spectrum in the order of *m/z*. If the baseline drift is considered mass-independent within the mass range, the baseline drift can be simplified by averaging the N individual lowest intensities, each of which is the lowest intensity for one of the N peaks of the mass spectrum. The baseline drift was determined in our approach by averaging the N lowest intensities of the entire spectrum while the N lowest intensities were selected through sorting the entire spectral data points in an ascending order. Therefore, among the N lowest intensities, instead of one from each peak, more than one might come from the data points of one peak, or none from another peak. However, the advantage of our approach for baseline drift determination is its simplicity with practically sufficient accuracy. Its simplicity results from the one-time sorting in an ascending order of the peak intensities that fishes out all the N lowest intensities simultaneously from the full data list. Its sufficient accuracy is due to that the average of a large number (N, generally in hundreds or more) of lowest intensities can very likely represent the trend of the baseline shift of the entire spectrum (Scheme 1B). Additional examples of the determined baseline drifts are demonstrated in the Supplementary Materials (Figures S1-S4).

Generation of the Accumulative Layer Thickness Curve from the Baseline Drift Corrected Mass Spectrum

Correction for noise levels was next conducted on the baseline drift-corrected mass spectrum. An identical procedure to the baseline drift deduction but with a number of n lowest intensities for deduction (as described in Experimental) was repeated on the baseline drift-corrected data points (i.e., Data Set 0), which yielded Data Set 1. The spectrum reconstructed from Data Set 1 could be visualized as a spectrum with one layer (Layer 1) wiped off from the bottom of the baseline drift-corrected spectrum. The thickness of Layer 1, TL_1, was the average of the n lowest intensities from Data Set 0. In general, Data Set i (1

≤ i ≤ m, m is the last Data Set containing ≥ n spectral data points) were generated by deduction of a layer (Layer i; having a thickness of TL_i by Equation 3) from Data Set i-1. When these layers (Layer 1, Layer 2, ..., Layer m) were laid along the y-axis of the spectrum, they were unequally distributed by varied thicknesses of individual layers (TL_1, TL_2, ..., TL_m, respectively) (Figure 1). Only small changes existed in the thickness of the first few deducted layers due to the dominant contribution of the low intense peaks to these layer deductions (see Experimental) (Figures 1B and 2A). When the low intensity peaks were just deducted completely from the Data Set, the following layer deduction then picked the data points from the remaining high intensity peaks present in the Data Set. This would subsequently result in a significant increase in the thickness of the layer which we defined as the transition layer (Figure 2A).

Before the transition, the thickness of each layer did not change significantly from layer to layer (Figure 1B) because the n lowest intensity data points in the current Data Set and in its previous Data Set were both from low intensity peaks that had small intensity differences between the neighboring data points (Scheme 1A). *After the transition*, the thickness of each layer increased significantly from layer to layer (Figure 1A) because the n lowest intensity data points in the current Data Set and in its previous Data Set were both from high intensity peaks that had big intensity changes (Scheme 1A). *During the transition*, the n lowest intensity data points in the current Data Set were, at least partially, from high intensity peaks while the n lowest intensity data points in its previous Data Set were exclusively from low intensity peaks. Therefore, the average of the n lowest intensities from the current Data Set increased significantly compared to that from the previous Data Set at the occurrence of the transition, and the rate of the increase at the transition should be differentiable from that before the transition (where the increase if any would be slight) and after the transition (where the increase would be dramatic) (Figure 2).

It was noted that the trends of these determined layer thicknesses were varied irregularly (Inset in Figure 2A). This irregulation made the curve regression difficult while a precise curve regression is essential to automatically locate the transition layer mathematically by an algorithm. We found that employing an accumulative layer thickness curve yielded from the TL_i data (i.e., a curve of ATL_i *vs.* layer i (Figure 2B), where ATL_i was calculated by Equation 4) successfully bypassed this difficulty.

Determination of the Transition Layer from the Accumulative Layer Thickness Curve

The accumulative layer thickness curve with the first layer data point eliminated was used for the curve fitting as described in Experimental (Figure 2B, Table 1). The curve was fitted by polynomial regression in the order of 6, by which all the examined mass spectra thus far were well fitted. Next, the derivatives of the regression equation were performed, including the first derivative to the fourth derivative. Among these derivatives, the first derivative represents the rate of change of the accumulative layer thickness with layer (i.e., $y' = d(ATL_x)/dx$); the second derivative represents the rate of change of the accumulative layer thickness change rate or the acceleration of the accumulative layer thickness with layer (i.e., $y'' = d(y')/dx$); and the third derivative represents the rate of change of the acceleration of the accumulative layer thickness with layer (i.e., $y''' = d(y'')/dx$). The roots of Equation 6 (i.e., the fourth derivative = 0) represent the layers that correspond to the extrema of the rate of change of the acceleration of the accumulative layer thickness (i.e., the extrema of the curve of y''' *vs*. x). Since the fourth derivative of a sixth order polynomial is in the order of 2, Equation 6 has two roots. We observed that the lower value of the two roots generally located within the first few layers which might indicate a type of transition of which we did not yet know the meaning while the higher value of the two roots best represents the transition layer.

It should be pointed out that although the sixth order polynomial regression followed by finding the zeros of the fourth derivative of this polynomial regression to determine the transition layer worked best thus far among the regressions we tested, we would leave the possibility open that the regression with any other formula to fit the curve might achieve a similar result or even more precise determination of the transition layer.

Determination of the Noise Level of a Mass Spectrum

Since polynomial regression on experimental data points is not a mechanism based modeling of data, it is necessary to assure that any potential uncertainty from single time polynomial regression is eliminated and the transition layer determined from the regression is stable. Ideally, if a regression equation that displays a precise fitting of the experimental data points could reflect a true understanding of the relationship underlying the data, this regression equation should be independent of the number of data points employed for the regression. In practice, a similar regression equation would be obtained when the same regression were performed on selected, less data points from the curve. Accordingly, we determined the transition layers from the narrowed accumulative layer thickness curves covering different numbers of data points from the curve, i.e., x = 1 to m-2, x = 1 to m-3, ..., and x = 1 to 7 (which is the minimal number of data points required for a sixth order polynomial regression). An example was tabulated (Table 1 where m = 24).

Intriguingly, we found that the determined transition layers and noise levels were minimally affected by reducing the number of the data points from the curve used for the regression if the transition layer was within the employed data points (the highlighted data in Table 1). These results indicated that the determination of the transition layer was independent of how many data points from the accumulative layer thickness curve were used for the regression analysis. It is important to implement these procedures to self check the stability and improve the liability of the determined transition layer. In addition, this stability validated the precise fitting of the accumulative layer thickness curve by the obtained polynomial regression equation and consequently the accurate determination of the transition layer from the regression. The accumulative layer thickness corresponding to each of the determined transition layers was calculated by Equation 8, and those that were consistent (within 30% deviation as described in Experimental) were averaged and designated as the noise level of the mass spectrum. A few examples of mass spectra demonstrating the baseline drifts and noise levels were provided in the Supplementary Materials (Figures S1-S4).

The Effects of the Step Length Setting on Determination of the Transition Layer

To further examine the stability of a determined transition layer, we generated a variety of accumulative layer thickness curves using different step lengths varied from 0.01 to 0.4, where the step length setting at 0.01 represents the finest process while the step length setting at 0.4 represents the roughest process for generating the curve (Figure 3) as described in Experimental. It is apparent that the bigger the step length is, the bigger the jump could be from one layer to the next. It is not surprising that a big jump may either overpass the transition layer or diminish the transition on the curve or both, which in turn may result in inaccurate determination of the transition layer. We found that if a step length (L) employed was in the range of ≥ 0.15, a very short transition was present or there was no transition at all; if the step length fell into the range of $0.04 < L < 0.15$, a short, but steady transition was present; and if the step length was ≤ 0.04, a long transition was present (Figure 3). It should be pointed out that a very long transition might not be beneficial since the transition might be smoothed out. Accordingly, in practice, we employed a step length in the range of $0.04 < L < 0.15$. Specifically, we employed $L = 0.07$ as a default setting in the program and routinely performed our data analysis with this step length.

We observed that the transition layer and corresponding noise level determined from each of the accumulative layer thickness curves generated by using step lengths in the range of 0.04 < L < 0.15 were minimally varied (Table 2). This also validated the accurate determination of the transition layer from the regression of the accumulative layer thickness curve whose generation was independent of the step length within a certain range (i.e., 0.04 < L < 0.15) (i.e., independent of how finely each layer deduction was processed).

Examples of Improved Reproducibility of Quantification under Varied Analyte Concentrations with Baseline Correction

Accurate quantification requires sufficient reproducibility under various alterations on experimental conditions (e.g., varied analyte concentrations due to differential analyte recovery, and varied chemical residues due to differential carry-over during sample preparation). Those alterations on experimental conditions, which are inevitable but may not be noticeable in most cases, can affect spectral baseline. To determine the effect of baseline correction on accurate quantification under varied analyte concentrations, we performed the baseline correction for a series of mass spectra acquired from sequentially diluted lipid solutions of a mouse myocardial lipid extract (Table 3 and Figure S5). The results indicate that the absolute noise levels are reduced at higher dilution rates likely due to the decrease in concentration of residual chemicals while the relative noise levels increase with dilution because of the decreased S/N from the increased proportion of baseline in peak intensity (Table 3). Importantly, it was demonstrated that the variation of the peak intensity ratio of a lipid species *vs*. the selected internal standard (i.e., ratiometric comparison) is substantially reduced with baseline correction (Table 3). This reduced variation represents improved reproducibility and therefore more accurate quantification. For example, the peak at *m/z* 758.6 (as indicated with arrows in Figure S5) represents a modestly low intense ion in the spectra. The variation of the ratio of its peak intensity relative to the internal standard at *m/z* 674.6 is reduced from 12.2% without baseline correction to 5.6% with baseline correction (Table 3). To those ions in much lower abundance, it is anticipated that baseline correction can improve the reproducibility of quantification even more dramatically due to the low S/N ratios of the low abundance peaks. These results further validate the powerful utility of our newly-developed baseline correction approach and clearly demonstrate the importance of baseline correction for accurate quantification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute on Aging/National Institute of Diabetes and Digestive and Kidney Diseases Grant R01 AG31675 (XH) and National Institutes of Health Grant P01 HL57278 (RWG). RWG and XH have financial relationships with LipoSpectrum LLC. RWG also has a financial relationship with Platomics, Inc.

References

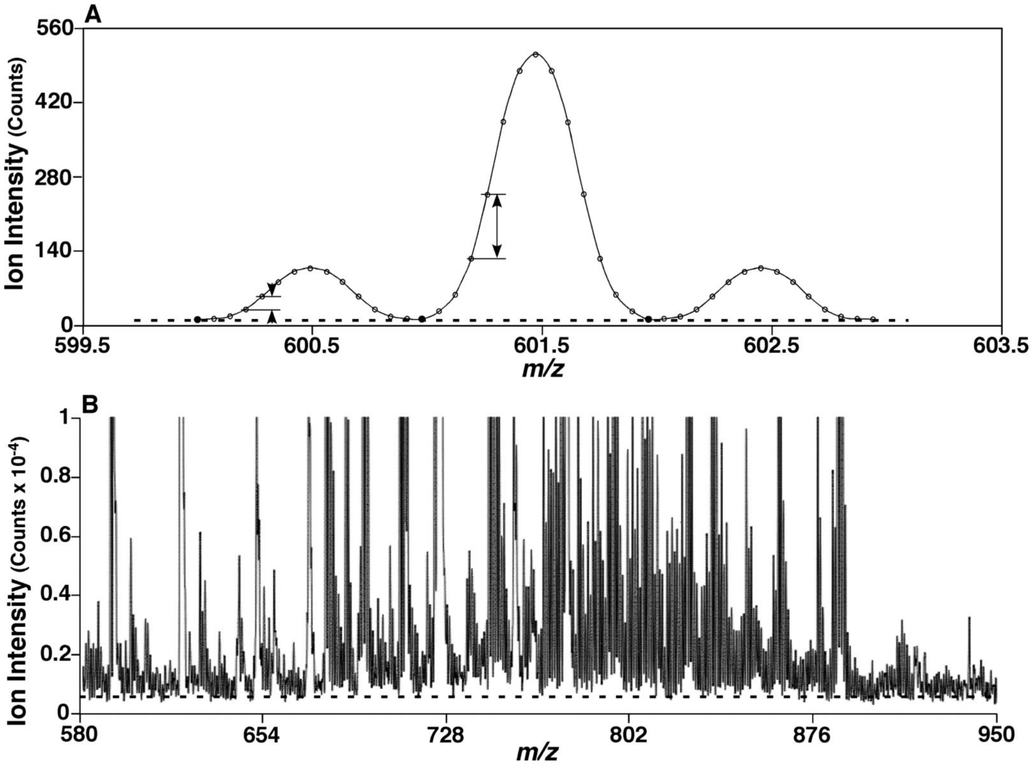
1. Han X, Gross RW. Shotgun Lipidomics: Electrospray Ionization Mass Spectrometric Analysis and Quantitation of the Cellular Lipidomes Directly from Crude Extracts of Biological Samples. Mass Spectrom. Rev. 2005; 24:367–412. [PubMed: 15389848]
2. Norris JL, Cornett DS, Mobley JA, Andersson M, Seeley EH, Chaurand P, Caprioli RM. Processing Maldi Mass Spectra to Improve Mass Spectral Direct Tissue Analysis. Int. J. Mass Spectrom. 2007; 260:212–221. [PubMed: 17541451]
3. Satten GA, Datta S, Moura H, Woolfitt AR, Carvalho Mda G, Carlone GM, De BK, Pavlopoulos A, Barr JR. Standardization and Denoising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens. Bioinformatics. 2004; 20:3128–3136. [PubMed: 15217815]

4. Ivanova PT, Milne SB, Byrne MO, Xiang Y, Brown HA. Glycerophospholipid Identification and Quantitation by Electrospray Ionization Mass Spectrometry. Methods Enzymol. 2007; 432:21–57. [PubMed: 17954212]

5. Han X, Yang J, Cheng H, Ye H, Gross RW. Towards Fingerprinting Cellular Lipidomes Directly from Biological Samples by Two-Dimensional Electrospray Ionization Mass Spectrometry. Anal. Biochem. 2004; 330:317–331. [PubMed: 15203339]

6. Han X, Yang K, Gross RW. Microfluidics-Based Electrospray Ionization Enhances Intrasource Separation of Lipid Classes and Extends Identification of Individual Molecular Species through Multi-Dimensional Mass Spectrometry: Development of an Automated High Throughput Platform for Shotgun Lipidomics. Rapid Commun. Mass Spectrom. 2008; 22:2115–2124. [PubMed: 18523984]

7. Williams, B.; Cornett, S.; Dawant, B.; Crecelium, A.; Bodenheimer, B.; Caprioli, RM. An Algorithm for Baseline Correction of Maldi Mass Spectra.. In: Kennesaw, G., editor. Proceedings of the 43rd Annual Southeast Regional Conference. Association for Computing Machinery; New York, NY, USA: 2005. p. 137-142.



Scheme 1.

Illustration of the baseline drift as well as intensity difference between neighboring data points. The solid symbols in a three-peak imitated mass spectrum (Panel A) indicate the 3 least intensity points which were averaged to determine the baseline drift. The arrows indicate the intensity differences between neighboring data points in either a low or high intensity peak (Panel A). The broken line in a mass spectrum acquired from a lipid extract (Panel B) indicates the determined levels of baseline drift.

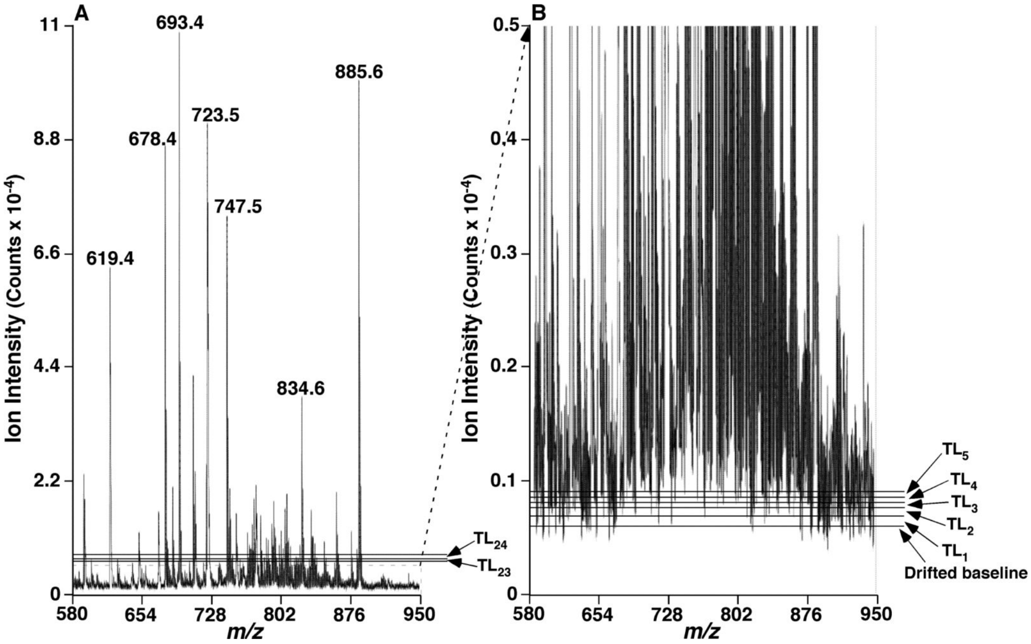


Figure 1.

Illustration of the layer thicknesses of a mass spectrum. Panel A displays a mass spectrum of mouse myocardial lipid extract acquired by negative-ion ESI-MS as described previously [6] and in supplementary materials. Panel B shows the amplified section of the mass spectrum displayed in panel A and illustrates the physical meaning of the baseline drift and the thickness of layer (TL). TL_1, TL_2, TL_3, TL_4, TL_5, TL_{23} and TL_{24} representatively indicate the thickness of layer corresponding to layers 1, 2, 3, 4, 5, 23, and 24 (the last layer), respectively.

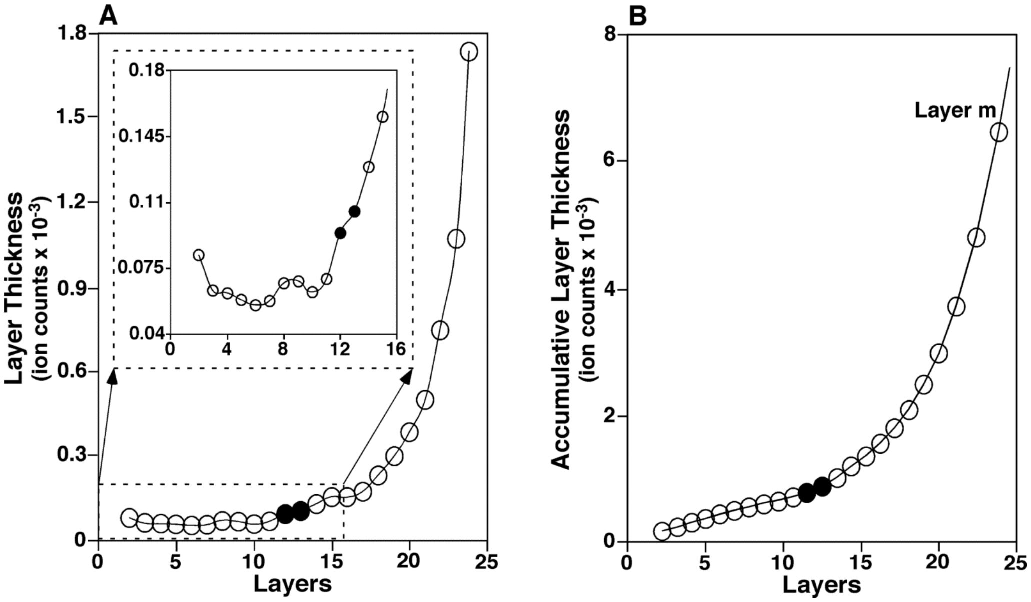


Figure 2.

Representative layer thickness curve and accumulative layer thickness curve. The layer thickness curve (Panel A) and accumulative layer thickness curve (Panel B) were derived from the mass spectrum in Figure 1A. The closed symbols indicate the region of the determined transition layer (12.1398, see Table 1).

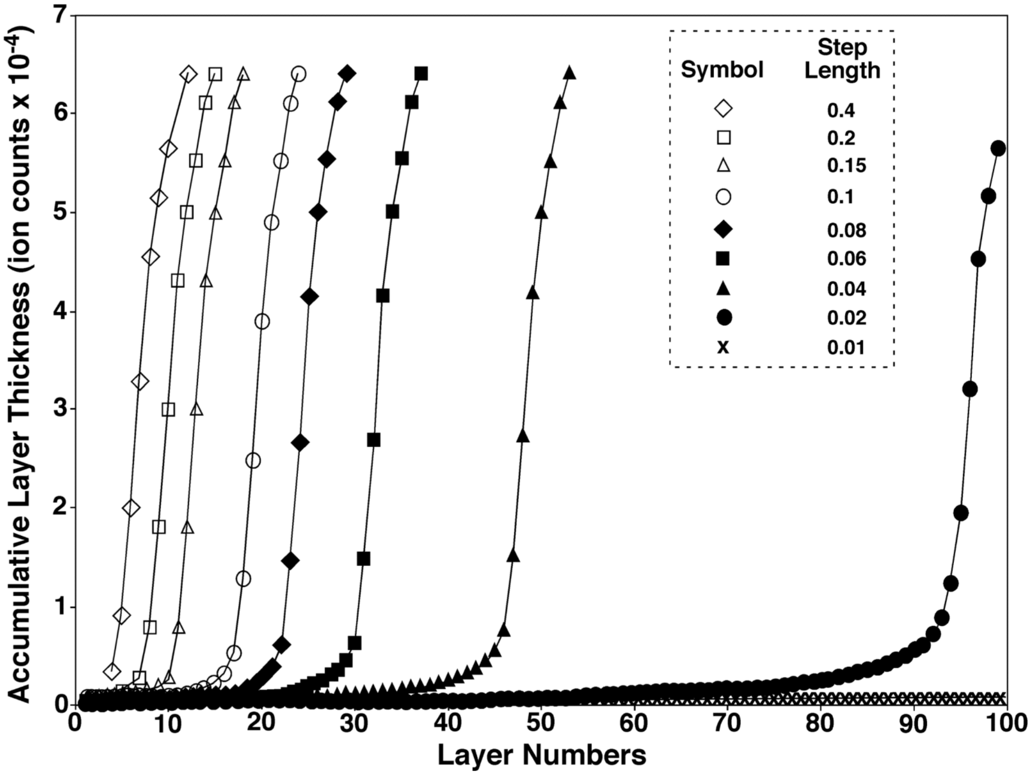


Figure 3.

The effects of the step length (L) on the generation of accumulative layer thickness curve of a mass spectrum. Each accumulative layer thickness curve corresponding to an individual step length ($0.01 \le L \le 0.4$) was constructed on the mass spectrum in Figure 1A as described under the section Experimental except the layer deduction procedure was repeated until no more spectral data points remained after deduction to generate an extended accumulative layer thickness curve for a clear demonstration purpose. The results indicate that if a large step length was employed (e.g., $L \ge 0.15$), a very short transition was present or there was no transition at all; if the modest step length was used (e.g., $0.04 < L < 0.15$), a short, but stable transition was present; and if the small step length was applied (e.g., $L \le 0.04$), a long transition was present. For clarification, only a few indicated accumulative layer thickness curves are displayed for demonstration.

Table 1

The effects of different numbers of layers used for determination of a transition layer to derive the noise level of a mass spectrum[a]

		The coefficients of the six order polynominal regression								
Layer i (x)	**ATL$_i$ or y**[b]	**x^6**	**x^5**	**x^4**	**x^3**	**x^2**	**x^1**	**x^0**	**Determined transition layer**	**Determined noise level**[b]
1[c]	96.0257									
2 (1)	178.1054									
3 (2)	241.4414									
4 (3)	303.5546									
5 (4)	362.1216									
6 (5)	417.7978									
7 (6)	475.7886									
8 (7)	543.0676	-0.0060	0.1401	-1.1303	3.9280	-6.7077	68.9562	112.9251	6.4240	442.3857
9 (8)	611.4270	-0.0129	0.3021	-2.6454	10.9921	-23.7935	88.9226	104.3376	6.1826	428.3870
10 (9)	673.6276	0.0015	-0.0781	1.2995	-9.2411	29.4259	22.3163	134.4206	14.3558	1126.6866
11 (10)	743.2630	0.0091	-0.3010	3.8417	-23.4735	69.9014	-31.7203	159.9766	8.0684	547.7460
12 (11)	836.9986	0.0064	-0.2158	2.7817	-17.0385	50.2164	-3.7860	146.1761	8.2008	556.7936
13 (12)	942.5386	0.0003	-0.0028	-0.0868	1.7348	-11.2756	88.6708	98.5609	7.1979	489.1044
14 (13)	1071.3711	-0.0009	0.0433	-0.7543	6.4208	-27.6442	114.6717	84.6279	**11.3401**	**775.1448**
15 (14)	1226.8370	-0.0009	0.0426	-0.7440	6.3441	-27.3593	114.1949	84.8934	**11.3739**	**778.3111**
16 (15)	1384.3546	-0.0013	0.0574	-0.9883	8.2921	-35.0162	127.6661	77.1149	**11.0511**	**748.0493**
17 (16)	1561.4457	-0.0007	0.0332	-0.5633	4.6994	-20.1054	100.1519	93.5721	**11.2062**	**762.5905**
18 (17)	1795.2227	0.0001	-0.0087	0.2179	-2.2787	10.3935	41.2455	130.0248	**12.5442**	**894.4364**
19 (18)	2097.3211	0.0005	-0.0285	0.6083	-5.9529	27.2633	7.2054	151.7936	**12.4177**	**881.0784**
20 (19)	2483.2300	0.0005	-0.0289	0.6166	-6.0358	27.6624	6.3657	152.3480	**12.4211**	**881.4445**
21 (20)	2983.0357	0.0004	-0.0229	0.4860	-4.6806	20.8400	21.3147	142.1702	**12.2109**	**859.2555**
22 (21)	3729.9181	0.0006	-0.0319	0.6910	-6.9057	32.5348	-5.3273	160.8592	**12.6271**	**903.1856**
23 (22)	4801.2543	0.0007	-0.0389	0.8581	-8.7999	42.9106	-29.8675	178.5809	**12.9036**	**932.3677**
24 (23)	6467.8738	0.0009	-0.0558	1.2786	-13.7695	71.2385	-99.3373	230.1834	**13.4421**	**999.4903**
								Average[d]	**12.1398**	**855.9413**

[a] The mass spectrum of mouse myocardial lipid extract acquired by negative-ion ESI-MS as described previously [6] was shown in Figure 1A and described in the supplementary materials. The accumulative layer thickness corresponding to each individual layer of the spectrum was calculated as described in the section Experimental and tabulated in the second column of the table. Different

numbers of layers (from layer 2 to layer 8, layer 2 to layer 9, . . ., layer 2 to layer 24 (Layer m, the last layer)) were used for determination of the transition layer through fitting their accumulative layer thicknesses with sixth order polynomial regression (see columns 3 to 9 of the table) followed by calculating the fourth derivative of the polynomial and solving the zero of the derivative as described in the section Experimental. Minimally varied transition layers (column 10) and corresponding noise levels (column 11) were determined from regression fitting data from layers 2 to 14, layers 2 to 15, . . .,layers 2 to 24 as highlighted in bold in the table.

[b] The unit is in ion counts.

[c] This first layer was not used for fitting (see Experimental).

[d] Average of the bold values.

Table 2

Examples of the constancy of the determined noise levels of mass spectra for lipid analysis as varied with the step length of the accumulative layer thickness curves

Step length	Determined baseline level (Example I) (Ion counts)	Determined baseline level (Example II) (Ion counts)	Determined baseline level (Example III) (Ion counts)
0.05	3527.7	1717.4	411164.1
0.06	3435.6	1622.2	404535.6
0.07	3350.6	1562.8	402887.7
0.08	3477.1	1555.6	402625.4
0.09	3358.2	1497.6	404002.9
0.10	3245.1	1542.4	401291.8
0.11	3234.3	1506.6	398201.2
0.12	3294.3	1505.9	398957.1
Mean ± SEM (Relative error)	3365.4 ± 38.0 (1.1%)	1563.8 ± 26.2 (1.7%)	402958.2 ± 1417.0 (0.4%)

Table 3

An example that baseline correction improves the accuracy of quantification[a]

Dilution factor[b]	Drift[c] (relative[d])	Noise[c] (relative[d])	Baseline level[c,e]	Base peak[c]	Ion peak at *m/z* 674 (I_{674})[c]	Ion peak at *m/z* 758 (I_{758})[c]	I_{758}/I_{674} (no baseline correction)	I_{758}/I_{674} (baseline corrected)
1	3.29 (0.60)	4.4 (0.81)	7.72	550	498.3	100.1	0.201	0.188
2	2.80 (0.91)	3.1 (1.00)	5.89	308	294.1	56.8	0.193	0.177
4	2.13 (1.09)	2.4 (1.24)	4.55	195	178.0	36.1	0.203	0.182
8	1.18 (1.38)	1.6 (1.94)	2.82	85	73.2	16.8	0.229	0.198
16	1.84 (2.87)	2.2 (3.44)	4.04	64	58.4	15.0	0.257	0.202
						Mean ± SD	0.217 ± 0.026	0.189 ± 0.011
						CV (%)	12.15	5.56

[a] The reduced variation of ion peak intensity ratios of I_{758}/I_{674} after baseline correction indicates the importance of baseline correction for accurate quantification. The ion peak at *m/z* 674 represents an internal standard and the ion peak at *m/z* 758 as indicated in Figure S5 represents a modestly low intense ion.

[b] A stock lipid extract of mouse myocardium was prepared by re-suspending dry lipid residue in 100 μL of solvent per mg of tissue protein followed by 100 fold dilution and was sequentially further diluted prior to direct infusion.

[c] Numbers are shown in absolute ion counts (**$\times 10^{-4}$**).

[d] The relative drift and noise (**$\times 10^{2}$**) are calculated by normalizing the absolute counts of drift and noise to the absolute counts of base peak and are given in parenthesis. The baseline level is the sum of drift and noise.