NIH Public Access
Author Manuscript
J Chromatogr B Analyt Technol Biomed Life Sci. Author manuscript; available in PMC 2010 September 15.

Published in final edited form as:
J Chromatogr B Analyt Technol Biomed Life Sci. 2009 September 15; 877(26): 2847–2854. doi:10.1016/j.jchromb.2008.12.043.

Algorithms for Automatic Processing of Data from Mass Spectrometric Analyses of Lipids

Haowei Song[1], **Jack Ladenson**[2], and **John Turk**[1,*]

[1]Mass Spectrometry Resource, Division of Endocrinology, Metabolism, and Lipid Research, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110

[2]Division of Laboratory and Genomic Medicine, Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110

Abstract

Lipidomics comprises large-scale studies of the structures, quantities, and functions of lipid molecular species. Recently developed mass spectrometric methods for lipid analyses, especially electrospray ionization (ESI) tandem mass spectrometry, permit identification and quantitation of an enormous variety of distinct lipid molecular species from small amounts of biological samples but generate a huge amount of experimental data within a brief interval. Processing such data sets so that comprehensible information is derived from them requires bioinformatics tools, and algorithms developed for proteomics and genomics have provided some strategies that can be directly adapted to lipidomics. The structural diversity and complexity of lipids, however, also requires the development and application of new algorithms and software tools that are specifically directed at processing data from lipid analyses. Several such tools are reviewed here, including LipidQA. This program employs searches of a fragment ion database constructed from acquired and theoretical spectra of a wide variety of lipid molecular species, and raw mass spectrometric data can be processed by the program to achieve identification and quantification of many distinct lipids in mixtures. Other approaches that are reviewed here include LIMSA (Lipid Mass Spectrum Analysis), SECD (Spectrum Extraction from Chromatographic Data), MPIS (Multiple Precursor Ion Scanning), FIDS (Fragment Ion Database Searching), LipidInspector, Lipid Profiler, FAAT (Fatty Acid Analysis Tool), and LIPID Arrays. Internet resources for lipid analyses are also summarized.

Keywords

Lipidomics; electrospray ionization (ESI); tandem mass spectrometry; algorithms and software tools

1. Introduction

Bioinformatics tools for automatic, computerized analyses of large data sets are essential components of the systematic and widely inclusive examination of genes and proteins now designated “genomics” and “proteomics”, respectively. There is a relative paucity of such tools

© 2008 Elsevier B.V. All rights reserved.

*To whom correspondence should be addressed at Washington University School of Medicine, Box 8127, 660 South Euclid Avenue, St. Louis, MO 63110; telephone 314-362-8190; FAX 314-362-7641, email jturk@DOM.wustl.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

available for comparable examinations of complex lipids ("lipidomics"), and it would be desirable to develop additional informatics tools for automatic identification and quantitation of lipid species from raw mass spectrometric data. Such data can now be obtained quickly from small amounts of biological materials and can contain sufficient information to profile hundreds to thousands of distinct molecular species [1]. The relatively recent application of electrospray ionization mass spectrometry (ESI/MS) to lipid analyses has resulted in much improved sensitivity and shorter analysis times than conventional methods [2,3].

Lipids comprise thousands of complex biomolecules with great structural diversity [4]. Conventional methods for lipid analyses have included two stage chromatographic separations of complex lipid mixtures into lipid classes, followed by a second analysis to isolate molecular species, and that process is laborious and time-consuming and involves analyte losses that limit sensitivity. The recently developed approach of "shotgun lipidomics" involves introduction of unfractionated lipid extracts into the ESI source of a tandem mass spectrometer and "intrasource separation" of molecular species based on the *m/z* values of the ions produced from them [1], followed by a second analytic dimension of tandem mass spectrometric profiling that permits identification and quantitation of hundreds of molecular species in a single sample [5]. HPLC separation methods coupled with ESI/MS(/MS) have also been described [3,6,7] and provide powerful technical platforms for identification and quantitation of large numbers of lipid molecular species in a high-throughput manner.

Such high throughput methods quickly generate huge amounts of data that are difficult to process manually. Data processing thus becomes a central and rate-limiting step in lipidomics, and this motivates development of computerized algorithms to process data from high-throughput ESI/MS analyses of lipids. In this review, we summarize algorithms and informatics tools that have been developed to process lipidomics data. The foundation of one such approach has been the systematic study of the ion chemistry of the fragmentation of complex lipid molecular species upon collisionally activated dissociation [8,9] to identify lipid structures coupled with high throughput database searching [9].

2. Workflow in Processing Data from Mass Spectrometric Analyses of Lipids

The data in acquired mass spectra are first converted to formats that permit processing of data from a variety of different types of mass spectrometers by available software tools. Lipid molecular species can then be identified by several approaches that include accurate mass determinations [10], two dimensional mass spectrometric fingerprinting [5], multiple precursor ion scanning [11], or MS/MS spectral database searching [9]. Once the lipid molecular species in a mixture are identified, precursor scan data can be processed to achieve quantification when internal standards have been included in the analyses of the mixture.

3. Converting and Processing Data Formats

Accurate and efficient data extraction from acquired mass spectra is the first step performed in all bioinformatics approaches to processing data from analyses of lipids or other molecules. Described tools include SECD and LIMSA, which process data by conversion to the NetCDF format using the DataBridge tool provided by MassLynx 4.0 software [12]. The LipidQA program [9] can directly process Finnigan *.RAW data. Data acquired from Waters mass spectrometers must first be processed by a program tool from MassLynx to extract the precursor ion and fragment ion data and generate *.pkl files. These files then contain *m/z* values of precursor ions and their intensities and a peak list of fragment ion *m/z* values and intensities. The LipidInspector program [13] processes dta files, which are peak list files for the SEQUEST [14] program and are generated by software tools from the instrument vendor. The Lipid Profiler program [11] can interface directly with Analyst software to process mass spectral

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

data. Most other programs, such as FAAT [10], transfer the raw mass spectral data to a peak list file in txt format for further analysis.

Almost every vendor of mass spectrometers has a proprietary data format, and it is impractical for software tools to support all such formats. In addition, processing extracted plain txt files (such as dta, pkl, or mgf) severely limits the ability of a program to process data further because such files fail to capture all of the originally acquired information, especially that required for quantitation. In an effort to construct a program for lipid analysis capable of quantitation that could process data from a variety of instruments from different vendors, Haimi and Somerharju [12] adopted a self-describing, machine-independent data format designated NetCDF. The mzXML [15] is a open data format for storage and exchange of mass spectrometric data for proteomics research that employs the more popular XML format, and it would be desirable to have an analogous open data format for lipidomics data that can be coupled with or include tools to convert vendor-specific format files for further processing.

4. Data Processing

The general data processing procedure is illustrated as Figure 1, and it includes baseline or noise reduction, smoothing, signal-to-noise ratio calculation, peak extraction, and deisotoping and deconvolution. The increasingly powerful computerized data systems of modern mass spectrometers permit extremely rapid sampling frequencies and scanning speeds, and recorded spectra contain thousands of data points. This allows weak signals to be detected when special data processing algorithms are used that avoid both overlooking and overprocessing peaks of low intensity. For example, inappropriate specification of symmetric windows and sub-intervals [16] can result in assignment of a Signal/Noise ratio for a peak of weak intensity that falls below the formal detection limits specified by the program. In the example in Figure 2, with a symmetric window of 100 Th, the calculated S/N value for the low intensity peak at *m/z* 762 is 5.15 if the spectrum is divided into sub-intervals of 5 Th. In contrast, if the selected sub-interval is 20 Th, the calculated S/N value is 2.81, which would fall below a specified threshold value of 3.0. Moreover, selection of an inappropriate smoothing width causes a reduction in peak amplitude and an increase in bandwidth, especially for spectra acquired at low sampling frequencies (Figure 3).

The first step in data processing is baseline or noise reduction. The programs SECD [12] and LipidQA [9] employ similar strategies to achieve noise reduction without obscuring weak signals. In general, these programs divide the spectrum into small segments, and the baseline noise level is estimated and minimized within individual segments.

Data smoothing algorithms must be employed carefully when it is desirable to identify weak signals and to quantitate the compounds that they represent. Using wide fields results in greater smoothing, but this is achieved at the potential cost of signal distortion by the smoothing operation. The optimal smoothing width depends upon peak shape and width and on the digitization interval. A critical parameter in avoiding spectrum distortion (Figure 3) for weak signals that contain few data points is the smoothing ratio, which is the ratio between the smoothing width and the number of points in the half-width of the peak. A smoothing ratio of less than 0.2 should be employed in order to avoid distortion of peak height and width [17]. In practice, if the same signal processing operations are applied to samples and to standards, the peak height reduction of the standard signals should be the same as that for the sample signals, and any distortions should cancel each other. If a greater smoothing ratio is required to improve efficiency for some applications, a combination of both internal and external standard correction methods can be employed, as in LipidQA [9]. This approach permits greater accuracy in the detection and quantitation of weak signals.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Deisotoping is another required step in data processing that accounts for [^{13}C] isotope effects and resolves overlapping isotopic peaks of lipid molecular species with ions of similar *m/z* values. A subtraction algorithm [18] is the most frequently employed deisotoping method in current approaches to analyzing lipidomics data[9,19,20], although other approaches have also been examined. Haimi and Somerharju compared three different deisotoping approaches that included algorithms for subtraction, linear fitting, and Gaussian peak model fitting, respectively [21]. Han and Gross reported a two-step method to correct for isotope effects [22]. In this approach, Factor Z1 is used to correct for the [^{13}C] effect arising from different numbers of carbon atoms in the target analyte and the internal standard, and Factor Z2 is used to correct for peak overlap between the second isotopic of one lipid molecular species with the monoisotopic peak from a second species represented by an ion with an *m/z* value 2 Th greater than that of the first species. This is a common circumstance in analyses of lipid mixtures, which often contain molecules that differ from each other by the presence or absence of a single double bond. Another algorithm developed by Ejsing and Shevchenko [11] for isotope correction involves summing the isotopic peak intensities of the monoisotopic ion with that of the first and second isotopomer obtained by precursor scanning.

5. Algorithms for Lipid Identification

5.a. Lipid Identification with Accurate Mass

FT-ICR and orbiTRAP mass spectrometers achieve measurement of *m/z* values with high resolution and mass accuracy, and this facilitates exclusion of some elemental compositions as candidates for species represented by an ion of a given observed *m/z* value. This has been exploited by Leavell and Leary in their program Fatty Acid Analysis Tool (FAAT) [10], which can process data from FT-ICR-MS analyses of lipids. Acquired data is subjected to scaling and reduction operations, and ions of given *m/z* values can be assigned to specific lipid molecular species by comparison with a user-defined library based on exact mass measurements. The high resolution and mass accuracy of FT-ICR-MS measurements also enables FAAT to process data from studies of the metabolism of stable isotope labeled precursors, and this is not possible for low resolution mass analyzers, such as quadrupoles, because of spectral overlap of nominally isobaric species of different elemental compositions. Other approaches permit assignment of the identities of lipid molecular species from *m/z* values measured on low resolution mass spectrometers when coupled with an HPLC separation step before mass analysis and a user-defined table that employs both retention time and *m/z* value for identification[12,19,23].

5.b. Lipid Identification with Neutral Loss or Product ion Scanning

Tandem quadrupole mass spectrometric scanning modes have also been employed to facilitate identification of multiple lipid species in complex mixtures. In neutral loss scanning, ions in the second mass analyzer are detected that differ in *m/z* value from that of their precursor ions by some specified, fixed value. In precursor ion scanning, parent ions from the first mass analysis step are identified that produce a fragment ion of a specified, fixed *m/z* value in the second mass analyzer. Such scans can be performed with great sensitivity and specificity on tandem quadrupole instruments, in part, because the method of data acquisition is altered in a way that maximizes signal and reduces noise. Approximations of such scans can be achieved by computational methods with data from other types of tandem mass spectrometers.

The approach of “shotgun lipidomics” developed by Han and Gross involves two dimensional mass spectrometric fingerprinting [1] in which lipid mixtures are analyzed by multiple neutral loss and precursor ion scans on a triple quadrupole mass spectrometer. Shevchenko and colleagues developed a multiple precursor ion scanning technique (MPIS) [24] for lipid identification from data acquired with an Applied Biosystems Inc. (ABI) hybrid quadrupole

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

time-of-flight (Q-tof) mass spectrometer, which can acquire multiple precursor ion spectra virtually simultaneously. In this approach, scans are performed in the first quadrupole with unit mass resolution, a 30 ms dwell time, and *m/z* scanning steps of 0.2 Th. At each step, a time-of-flight scan is then performed with dynamic collision energy of 45 eV to 60 eV. This approach is employed in the program Lipid Profiler [11], which processes virtually simultaneously 41 precursor negative ion spectra for a set of *m/z* values that include those of carboxylate anions of fatty acid substituents commonly encountered in complex lipids as well as ions that are lipid head-group-specific or other class-specific fragment ions. About 200 molecular species in a biological lipid mixture can be recognized Lipid Profiler [11].

Precursor ion scanning in negative ion mode fails to identify lipid classes that produce weak or absent class-specific ions upon CAD of the parent ion. Such classes include phosphatidylserine (PS), triacylglycerol (TAG), and sphingomyelin. The program Lipid Inspector [13] was developed by Shevchenko and colleagues to address this problem. In this approach, MS/MS data are acquired in a data-dependent manner and then exported as dta files. Lipid Inspector identifies product ions that differ in *m/z* value by a specified, fixed value from that of the precursor ion, thus approximating a neutral loss scan. Like Lipid Profiler, Lipid Inspector also identifies parent ions that yield fragment ions of specified *m/z* values, thus approximating a precursor ion scan. The resultant neutral loss and precursor ion scanning data are then processed to identify a wide range of lipid molecular species.

5.c. Lipid Identification with MS/MS Spectral Database Searching

Searching amino acid sequence databases is used to identify peptides and the proteins from which they are derived in proteomics studies. These observed tandem mass spectra are compared to a set of theoretical fragment ion spectra from all possible protein sequences. This is feasible for identification of peptides and proteins, which contain various arrangements and combinations of 20 amino acids. The problem is much more difficult for lipids because of the great structural diversity of their component substituents. Studies of the ion chemistry of the fragmentation upon CAD(collisionally activated dissociation) of lipid molecular species [8, 25–36] have resulted in the generation of a set of fragmentation rules that predict a theoretical product ion spectrum for a wide range of structurally diverse lipid molecular species. Figure 4 contains the tandem spectra of the isomeric GPC lipid molecular species 16:0/18:1-GPC and 18:1/16:0-GPC as Li^+ adducts. Table 1 contains a list of the fragment ions used to identify the head group and fatty acid substituents. As indicated in the table, the relative intensities of ions arising from losses of the fatty acid substituents reflects their position on the glycerol backbone.

Such a fragment ion database has been constructed from calculated and acquired reference tandem spectra of glycerophospholipids of all major head-group classes and incorporated into the LipidQA program [9] for identification and quantitation of lipid molecular species from raw mass spectrometric data. In this program, an acquired tandem mass spectrum is compared to the set of theoretical fragment ion spectra in the database to identify the lipid molecular species represented by the parent ion from which the tandem spectrum was obtained [9]. Each acquired tandem spectrum is iteratively compared to the fragment ion database to generate a set of candidate lipid molecular species that might correspond to the species represented by the parent ion. This set includes only those molecular species represented by an ion with an *m/z* value identical to that of the observed parent ion in question in the acquired spectrum. In this usage, “identical” means that the *m/z* values of the observed and candidate ions differ by no more than a specified value that is affected by the resolving power and mass accuracy of the instrument on which the data were acquired. Theoretical tandem spectra for this set of candidate compounds are then compared to the acquired spectrum to determine the best match.

A score is calculated on the basis of this comparison that reflects how well each theoretical spectrum matches the acquired spectrum. The scoring method used in the LipidQA program

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

involves determination of the number of fragment ions in the acquired spectrum that match ions in the theoretical spectrum. The ratio of that value and the total number of spectral lines in the theoretical tandem spectrum is then calculated. More recently, a probability-based scoring method has been developed in which the score is assigned the he $-10\log(P)$, where P is the statistical probability that the best match of the acquired and candidate spectra results from a random event. P is calculated as $\left(\left(\frac{1}{N_{\text{reference}}} \times \frac{1}{N_{\text{acquired}}}\right)^{N_{\text{match}}}\right)$ where N represents the number of spectral lines. Fragment ion database searching (FIDS) considers all fragment ions in the tandem spectrum, and therefore is in principle capable of more accurately identifying a wider range of lipid molecular structures than are scanning-based methods that involve comparisons only of selected ions. FIDS can be performed with a variety of different types of mass spectrometers and does not require *m/z* value data acquired on an instrument with high resolution and mass accuracy, although high mass accuracy does facilitate data processing by reducing the number of candidate matches. FIDS can be applied to MS data acquired by intrasource separation or from on-line HPLC/MS approaches.

6. Interference

Isobaric interference is a common problem in lipid analysis even when sample pretreatment or separation steps are performed before mass analysis. In addition, fatty acids, especially polyunsaturates, can yield fragment ions upon CAD that can be confused with other fatty acid substituents that are not contained in the molecular species being analyzed. These factors complicate identification of lipid species by neutral loss or precursor ion scanning methods that consider only selected features of the tandem mass spectrum [11]. The FIDS algorithm of the LipidQA program [9] achieves more reliable identification and is in principle capable of identifying a wider range of structurally distinct lipid molecular species because it considers data from the entire tandem mass spectrum.

An example that illustrates this point is provided by our experience in examining the tandem spectrum produced from CAD of a parent ion of *m/z* 865 in negative ion mode from ESI/MS analysis of a mouse peritoneal leukocyte lipid extract. Based on the acquired tandem spectrum (Figure 5), LipidQA identified the parent ions as $[M\text{-}H]^-$ of 22:6/22:6-GPG with an ID score of 0.77. The tandem spectrum, however, contains an intense ion at *m/z* 283, which is isobaric with stearate (18:0) anion. A precursor of *m/z* 283 scanning-based identification algorithm thus might confuse 22:6/22:6-GPG with the isobaric species 18:0/18:0-GPI. LipidQA assigned a low ID score of 0.12 for the comparison of the acquired spectrum with the database spectrum of 18:0/18:0-GPI because of the absence from the acquired spectrum of several prominent ions in the database spectrum of 18:0/18:0-GPI.

7. Quantitation

Once the lipid molecular species in a sample have been identified, it is often desirable to determine their quantities. Such quantitative data permits comparisons of differences in amounts of a given molecular species in treated and control samples after an experimental biological perturbation, for example, or between normal and diseased tissue or biological fluids. Approaches to obtaining quantitative MS data include metabolic stable isotope labeling [10] or addition of stable isotope labeled internal standards [37], *inter alia*. Sufficient resolution is required to discriminate among potentially overlapping peaks representing isobaric but distinct substances in complex biological mixtures.

Accurate quantitation with low resolution mass spectrometers, such as quadrupoles, has been demonstrated for multiple lipid molecular species in biological extracts in studies involving addition of one or two internal standards, especially after correction for [^{13}C] isotope effects

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and for variations in ionization efficiency that arise from differences in the number of carbon atoms and double bonds among distinct molecular species in a mixture [22]. The LipidQA program employs a single internal standard for each lipid class in the mixture. This permits normalization of the signal from different molecular species within each lipid class. Multiple standard curves are used to correct for differences in the number of carbon atoms between the target analyte and the internal standard and to correct for effects on signal intensity that arise from differences in degree of unsaturation. The set of calibration samples for each lipid class includes standard compounds that represent the species most often observed in biological extracts. The standard curve is determined by ESI/MS analyses of a series of samples prepared with a constant amount of internal standard and varied amounts of the target analyte. The signal intensity of each standard lipid species is normalized to that of the internal standard, and the normalized signals are used to generate a regression line versus the concentration of each standard. In the analysis of biological samples, the same amount internal standard used to prepare the calibration curves is added, and the signal for each identified lipid molecular species is normalized to that of the internal standard. The normalized value is then compared to the regression line of the calibration curve to determine the amount of that species contained in the sample. Figure 6 illustrates the general diagram of calibration method [9].

The Lipid Array program [38] employs a novel method for semi-quantitation without internal standards that is based on two central concepts. First, data normalization and Shewhart control charts are used to determine whether or the difference between two samples is random. To compensate for variations in peak intensities that arise from sample preparation steps and ionization efficiency, a unitless measure of the overall pattern of the observed spectrum is generated to compare changes in amounts of lipid molecular species at the cellular level. Two methods have been tested for data normalization. The first normalizes peak intensity to the mean and standard deviation of intensities observed at all *m/z* values in the spectrum using the equation $I^* = (I\text{-mean})/SD$. The second normalization method compares the rank of the signal intensity of a given peak to that of all other peaks in the data set. A Shewhart control chart is then constructed for every peak with the normalized intensity signal means, and control limits are established if the baseline signal is stable over time.

Within these limits, the program then examines the means of the normalized data from the stimulated condition for non-random variation. It is then determined whether there are statistically significant differences between the experimental and control conditions for each [(time point) *vs*. (peak *m/z* value)] pair. The results are then grouped and displayed as a comprehensive array with the *m/z* values on the vertical axis and time points on the horizontal axis. If the [(time point) *vs*. (peak *m/z* value)] regression is found to be increasing, a positive score is assigned, and if it is decreasing, a negative score is assigned. If there is no statistically significant change, a score of zero is assigned. An underlying assumption in the Lipid Array analysis approach is that the total amount of lipid and the majority of molecular species are not different in the experimental and control samples.

8. Validation of Identification and Quantitation

Recently developed techniques for acquiring lipidomics data and bioinformatics tools to process them permit presumptive identification and quantitation of a huge number of lipid molecular species from many samples within a brief interval. There are relatively few formal demonstrations that these methods correctly identify the lipid molecular species actually contained in the analyzed mixture and accurately determine their quantities. A probability-based scoring method has recently been incorporated into the LipidQA [9] program that is aimed at statistical validation of lipid molecular species identification. With respect to quantification, the enhanced Lipid Array analysis [39] program introduces the concept of a significance score, which is the number of experiments that are statistically distinguishable

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

from the basal or control condition (established from 10 replicate trials) at a given time point. In this approach, a significance score greater than 4 or less than −4 is considered to represent statistically a significant difference between the control and experimental conditions.

9. Interpreting Results from Lipidomics Data-Processing Algorithms and Transferring Biological Information

Once lipidomics data processing is complete, the results must be interpreted to determine what information they provide about the issues or questions under examination in the study from which the analyzed samples were derived, and such interpretation is often facilitated by easily comprehensible data displays. Lipidomics software tools should provide a complete display of identified lipids and their quantities for an entire experiment or study. Significant information may reside in the list of identified lipid molecular species or their abundances or in comparisons of lipidomics data with genomics or proteomics information. Results from data-processing algorithms can be integrated into recognized signaling pathways by searching the Kyoto Encyclopedia of Genes and Genomes (KEGG) [40] pathway maps database, which includes a section that represents current information on lipid metabolic molecular interactions and reaction networks. The KEGG Brite (http://www.genome.ad.jp/kegg/brite.html) is a collection of hierarchical classifications that represent current information on various aspects of biological systems. Additional information is also provided, including genomic and molecular data for inferring higher order function analyses of lipid functions. SphinGOMAP is a category-specific lipid database that contains pathway maps for about 400 distinct sphingolipid and glycosphingolipid species [41]. LIPID MAPS Biopathways Workbench is a graphic tool that permits display, editing, and analyses of lipid pathways [42].

10. Internet Resources for Lipid Analyses

Lipid library (http://www.lipidlibrary.co.uk/) is a website that provides an extensive collection of information about lipids, including definitions of terms encountered in studies of lipids as well as descriptions of the structures, composition, occurrence, biochemistry, and functions of lipids. The site also provides practical and theoretical descriptions of a variety of experimental techniques for lipid analysis.

LipidMaps (http://www.lipidmaps.org/) provides a comprehensive lipid classification and nomenclature system, a database of lipid structures, and software tools to illustrate lipid pathways and analyses.

LipidBank (http://www.lipidbank.jp/) contains descriptions of more than 7000 lipid species, and experimental data, including mass spectra, are provided for some of them.

CyberLipid (http://www.cyberlipid.org/) contains descriptions of a large variety of lipid classes and includes references and protocols for experimental analyses of lipids.

11. Challenges and Perspectives

Lipidomics is constrained by bioinformatics bottlenecks that are also encountered in genomics and proteomics. Recently developed mass spectrometric methods for lipid analysis provide sufficient information to identify and measure an enormous number of distinct lipid molecular species and can generate huge amounts of experimental data within a brief interval. Processing such data so that comprehensible information is derived from them requires bioinformatics tools, and algorithms developed for proteomics and genomics have provided some strategies that can be directly adapted to lipidomics. The structural diversity and complexity of lipids, however, also requires the development and application of new algorithms and software tools that are specifically directed at processing data from lipid analyses. A variety of such software

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

tools to process lipidomics data are reviewed here, although additional tools may exist or be under development.

Aspects of lipidomics software development that deserve further attention include:

1. Open XML-based data formats need to be established that are widely accepted within the lipidomics research community for conversion of data in various formats acquired on mass spectrometers of different types from different vendors. Tools for transferring such data also need to be developed.
2. Methods for automated validation of algorithms and software tools for lipid identification and quantification require further development.
3. Additional tools that facilitate comprehension and interpretation of the large amount of information contained even in processed lipidomics data are needed, as are additional tools to relate lipidomics information into biological functional and interaction networks.

ACKNOWLEDGMENTS

Work in the authors' laboratories was supported by United States Public Health Service Grants R37-DK34388, P41-RR00954, P60-DK20579, and P30-DK56341. The authors thank Dr. Fong-Fu Hsu for many helpful discussions and Alan Bohrer for excellent technical assistance.

REFERENCES

1. Han X, Gross RW. Mass Spectrom Rev 2005;24:367. [PubMed: 15389848]
2. Han X, Gross RW. Proc. Natl. Acad. Sci. U.S.A 1994;91:10635. [PubMed: 7938005]
3. Kim H-Y, Wang T-CL, Ma Y-C. Anal. Chem 1994;66:3977. [PubMed: 7810900]
4. Murphy RC, Fiedler J, Hevko J. Chem. Rev 2001;101:479. [PubMed: 11712255]
5. Han X, Gross RW. J. Lipid Res 2003;44:1071. [PubMed: 12671038]
6. Vernooij EAAM, Brouwers JFHM, Bosch JJK-Vd, Crommelin DJA. J. Sep. Sci 2002;25:285.
7. Hvattum E, Hagelin G, Larsen Å. Rapid Comm. Mass Spectrom 1998;12:1405.
8. Hsu, FF.; Turk, J. Electrospray ionization with low-energy collisionally activated dissociation tandem mass spectrometry of complex lipids: structural characterization and mechanisms of fragmentation. Champaign, IL: AOCS publication; 2005.
9. Song H, Hsu F-F, Ladenson J, Turk J. J. Am. Soc. Mass Spectrom 2007;18:1848. [PubMed: 17720531]
10. Leavell MD, Leary JA. Anal. Chem 2006;78:5497. [PubMed: 16878888]
11. Ejsing CS, Duchoslav E, Sampaio J, Simons K, Bonner R, Thiele C, Ekroos K, Shevchenko A. Anal. Chem 2006;78:6202. [PubMed: 16944903]
12. Haimi P, Uphoff A, Hermansson M, Somerharju P. Anal. Chem 2006;78:8324. [PubMed: 17165823]
13. Schwudke D, Oegema J, Burton L, Entchev E, Hannich JT, Ejsing CS, Kurzchalia T, Shevchenko A. Anal. Chem 2006;78:585. [PubMed: 16408944]
14. Eng JK, McCormack AL, Yates JR. J. Am. Soc. Mass Spectrom 1994;5:976.
15. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R. Nat. Biotech 2004;22:1459.
16. Samuelsson J, Dalevi D, Levander F, Rognvaldsson T. Bioinformatics 2004;20:3628. [PubMed: 15297302]
17. O'Haver, T. 2008. http://www.wam.umd.edu/~toh/spectrum/Smoothing.html
18. Horn DM, Zubarev RA, McLafferty FW. J. Am. Soc. Mass Spectrom 2000;11:320. [PubMed: 10757168]
19. Kurvinen JP, Aaltonen J, Kuksis A, Kallio H. Rapid Comm. Mass Spectrom 2002;16:1812.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

20. Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. Biochim. Biophys. Acta (BBA) - Molecular and Cell Biology of Lipids 2004;1686:108.
21. Meija J, Caruso JA. J. Am. Soc. Mass Spectrom 2004;15:654. [PubMed: 15121194]
22. Han X, Gross RW. Anal. Biochem 2001;295:88. [PubMed: 11476549]
23. Hermansson M, Uphoff A, Kakela R, Somerharju P. Anal. Chem 2005;77:2166. [PubMed: 15801751]
24. Ekroos K, Chernushevich IV, Simons K, Shevchenko A. Anal. Chem 2002;74:941. [PubMed: 11924996]
25. Hsu F-F, Bohrer A, Turk J. J. Am. Soc. Mass Spectrom 1998;9:516. [PubMed: 9879366]
26. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 1999;10:587. [PubMed: 10384723]
27. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 1999;10:600. [PubMed: 10384724]
28. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2000;11:986. [PubMed: 11073262]
29. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2000;11:892. [PubMed: 11014451]
30. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2000;11:797. [PubMed: 10976887]
31. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2001;12:1036.
32. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2003;14:352. [PubMed: 12686482]
33. Hsu F-F, Turk J. J. Am. Soc. Mass Spectrom 2005;16:1510. [PubMed: 16023863]
34. Hsu F-F, Turk J, Shi Y, Groisman EA. J. Am. Soc. Mass Spectrom 2004;15:1. [PubMed: 14698549]
35. Hsu F-F, Turk J, Thukkani AK, Messner MC, Wildsmith KR, Ford DA. J. Mass Spectrom 2003;38:752. [PubMed: 12898655]
36. Hsu, F-Fu; Turk, J. J. Am. Soc. Mass Spectrom 2003;14:352. [PubMed: 12686482]
37. Murphy RC, James PF, McAnoy AM, Krank J, Duchoslav E, Barkley RM. Anal. Biochem 2007;366:59. [PubMed: 17442253]
38. Ivanova PT, Milne SB, Forrester JS, Brown HA. Mol. Interv 2004;4:86. [PubMed: 15087482]
39. Forrester JS, Milne SB, Ivanova PT, Brown HA. Mol. Pharmacol 2004;65:813. [PubMed: 15044609]
40. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. Nucl. Acids Res 2004;32:D277. [PubMed: 14681412]
41. Sullards MC, Wang E, Peng Q, Merril AH. Cell Mol. Biol 2003;49:789. [PubMed: 14528916]
42. Fahy E, Sud M, Cotter D, Subramaniam S. Nucl. Acids Res 2007;35:W606. [PubMed: 17584797]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

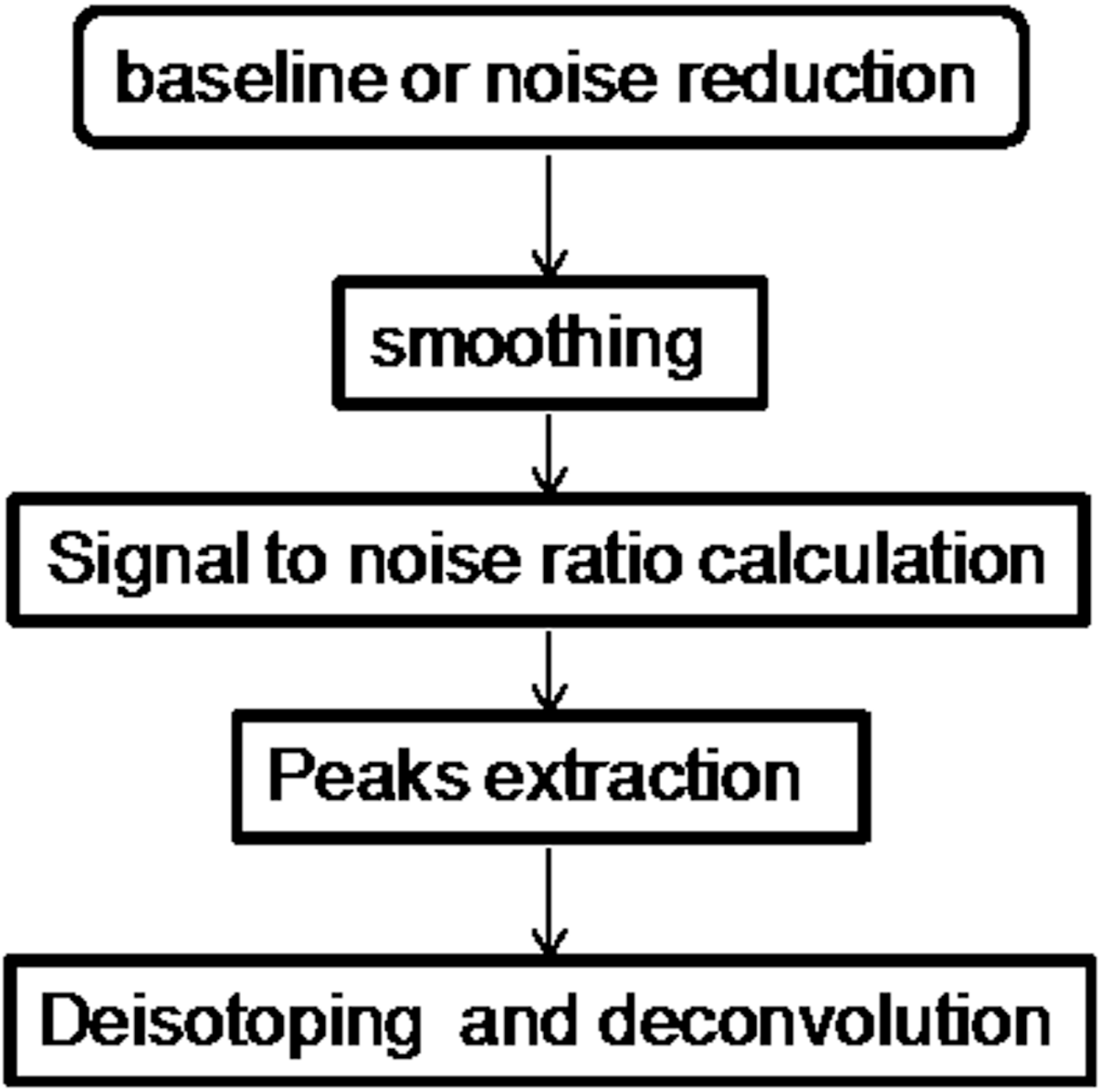

Figure 1.
General procedure for data processing

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

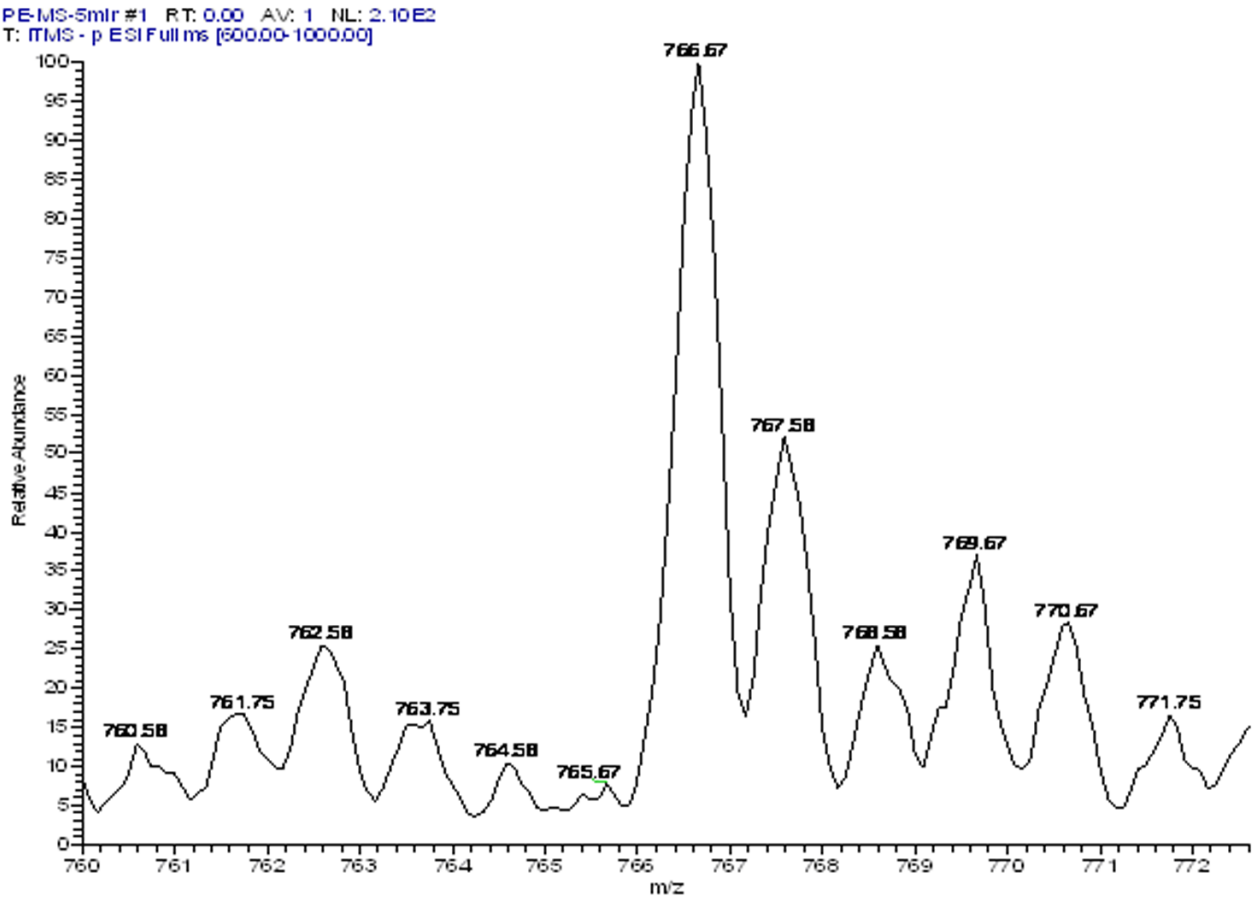

Figure 2.

MS spectrum of 18:2/20:4 GPE and 18:0/20:4 GPE. With a symmetric window of 100 Th, the calculated S/N value for the low intensity peak at m/z 762 is 5.15 if the spectrum is divided into sub-intervals of 5 Th. In contrast, if the selected sub-interval is 20 Th, the calculated S/N value is 2.81, which would fall below a specified threshold value of 3.0.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

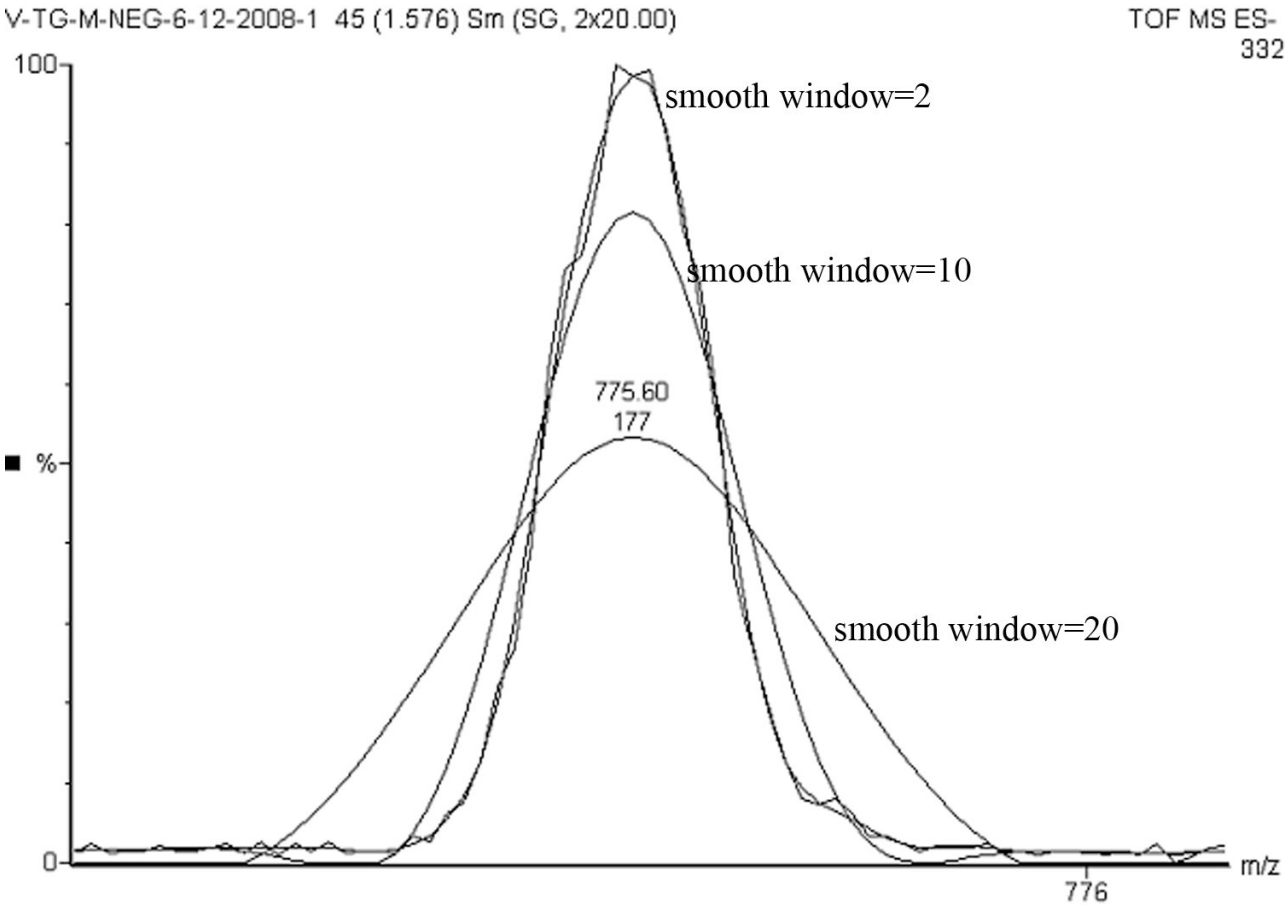

Figure 3.
Selection of an inappropriate smoothing width causes a reduction in peak amplitude and an increase in bandwidth.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

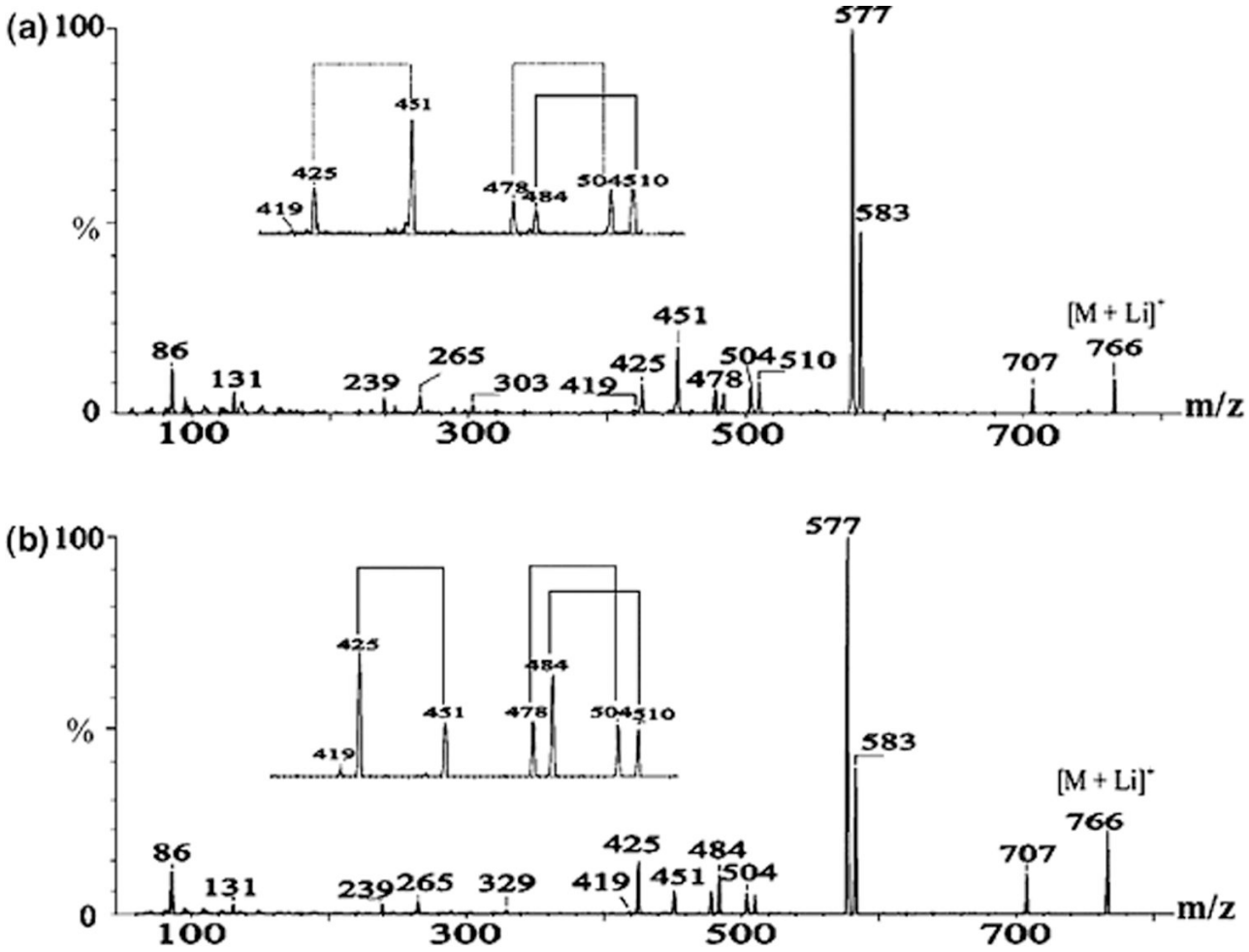

Figure 4.

MS/MS spectra of 16:0/18:1-GPC-Li^+ (**a**) and 18:1/16:0-GPC-Li^+ (**b**) obtained with a Finnigan TSQ-7000 tandem quadrupole instrument via product ion scan mode [9].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

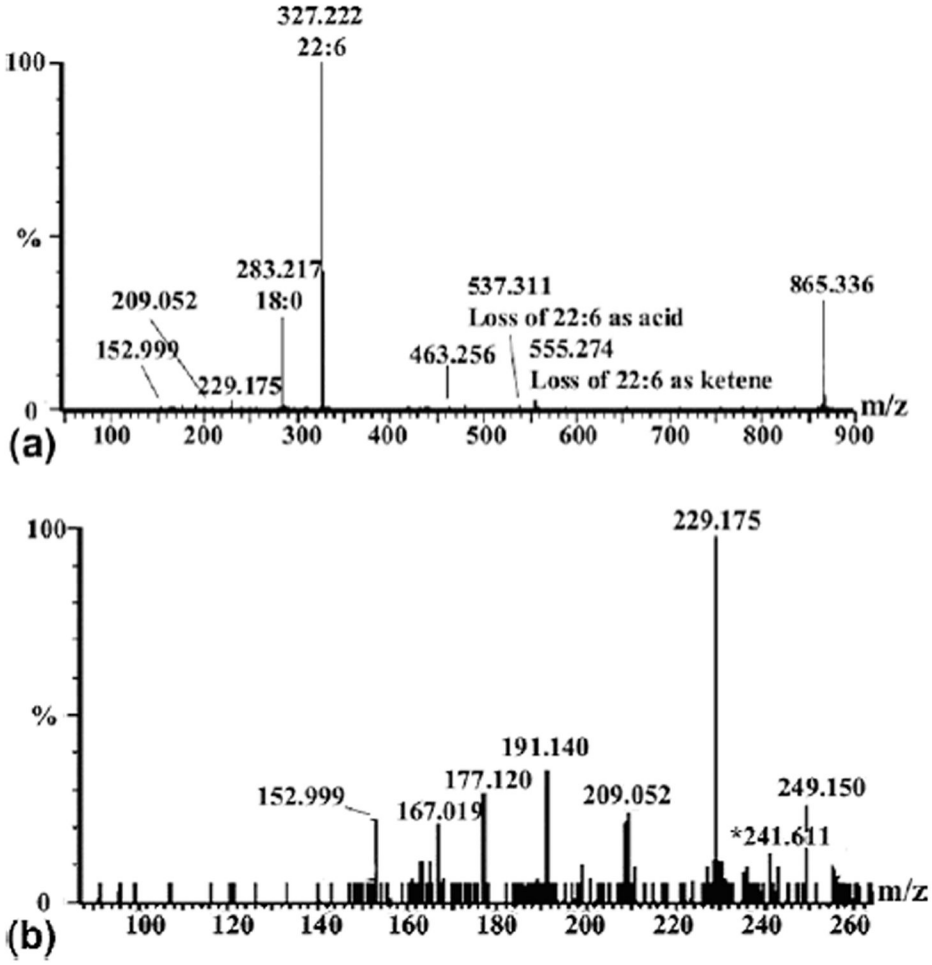

Figure 5.

MS/MS spectrum obtained from CAD of the ion of m/z 865.33 in a lipid extract from mouse peritoneal leukocytes. (**a**) Complete spectrum. (**b**) Expanded spectrum from m/z 80 to 260. Data were acquired with a Micromass Q-TOF Micro mass spectrometer in negative ion mode [9].

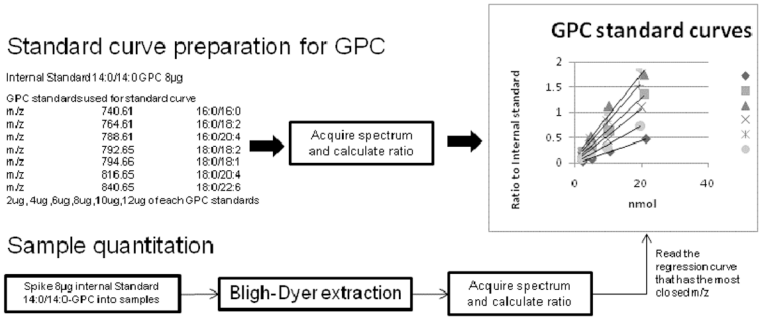

Figure 6.
Calibration method used in the LipidQA program involves a combination of corrections from both internal and external standards.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

J Chromatogr B Analyt Technol Biomed Life Sci. Author manuscript; available in PMC 2010 September 15.

Table 1

Examples of rules for fragmentation and relative intensities used to construct the fragment ion database [9].

Fragment Ions	Structural Feature Reflected[a]	Substituent Position	16:0/18:1-GPC-Li^+		18:1/16:0-GPC-Li^+	
			Fragment ion *m/z*	Observed Relative Intensity (%)	Fragment ion *m/z*	Observed Relative Intensity (%)
$[M+Li]^+$	NN	−	766.58	9	766.58	22
$[M+Li-59]^+$	HG	−	707.51	7	707.51	10
$[M+Li-183]^+$	HG	−	583.51	47	583.51	39
$[M+Li-189]^+$	HG	−	577.51	100	577.51	100
$[M+Li-R_1CO_2H]^+$	**R_1**	**+**	**510.34**	**8**	**484.32**	**10**
$[M+Li-R_2CO_2H]^+$	**R_2**	**+**	**484.32**	**5**	**510.34**	**5**
$[M+Li-59-R_1CO_2H]^+$	**R_1**	**+**	**451.27**	**17**	**425.25**	**14**
$[M+Li-59-R_2CO_2H]^+$	**R_2**	**+**	**425.25**	**7**	**451.27**	**7**
$[M+Li-R_1CO_2Li]^+$	**R_1**	**+**	**504.33**	**7**	**478.32**	**6**
$[M+Li-R_2CO_2Li]^+$	**R_2**	**+**	**478.32**	**6**	**504.33**	**5**
$[R_1CO]^+$	R_1	−	239.24	4	265.25	4
$[R_2CO]^+$	R_2	−	265.25	5	239.24	3
$[C_2H_5PO_4+Li]^+$	HG	−	131.01	6	131.01	3
$[C_5H_{12}N]^+$	HG	−	86.1	12	86.1	11

[a] the indicated ions contain information about the following structural features of the phospholipids molecule: NN (Number of Nitrogen atoms); HG (head group); R_1 (*sn*-1 substituent); R_2 (*sn*-2 substituent); substituent position refers to *sn*-1 vs. *sn*-2 glycerol carbon atom. Positional assignments: $[M+Li-R_1CO_2H]^+ > [M+Li-R_2CO_2H]^+$, $[M+Li-59-R_1CO_2H]^+ > [M+Li-59-R_2CO_2H]^+$, $[M+Li-R_1CO_2Li]^+ > [M+Li-R_2CO_2Li]^+$.