Bioinformatics tools and challenges in structural analysis of lipidomics MS/MS data

Jürgen Hartler, Ravi Tharakan, Harald C. Köffeler, David R. Graham and Gerhard G. Thallinger

Submitted: 29th February 2012; Received (in revised form): 16th May 2012

Abstract

Lipidomics, the systematic study of the lipid composition of a cell or tissue, is an invaluable complement to knowledge gained by genomics and proteomics research. Mass spectrometry provides a means to detect hundreds of lipids in parallel, and this includes low abundance species of lipids. Nevertheless, frequently occurring isobaric and isomeric lipid species complicate lipidomics analyses from an analytical and bioinformatics perspective. Various MS/MS strategies have evolved to resolve ambiguous identifications of lipid species, and these strategies have been supported by corresponding bioinformatics analysis tools. This review intends to familiarize readers with available bioinformatics MS/MS analysis tools and databases, the structural information obtainable from these, and their applicability to different MS/MS strategies. Finally, future challenges in detecting double bond positions are investigated from a bioinformatics perspective.

Keywords: lipidomics; MS/MS data analysis; high-throughput; MS/MS detection algorithms; analysis tools

INTRODUCTION

The high-throughput analysis of diverse biomolecules has provided fundamental insights into complex mechanisms in life and disease. These ‘omics’ research fields are characterized by large-scale analyses of various biological materials in a high-throughput manner. Lipidomics, a relatively young sub-discipline of metabolomics, explores the whole lipid content of cells, tissues and organisms [1, 2]. This scientific field has become of major interest recently, because its findings provide knowledge complementary to established disciplines such as transcriptomics and proteomics. Novel insights can be gained because lipids are products of highly regulated metabolic reaction pathways [3]. Recent research indicates the important role of lipids in major diseases, such as diabetes mellitus [4], Alzheimer’s disease [5, 6], cancer [7, 8], multiple sclerosis [9] and schizophrenia [10]. The effects of lipids in nutrition and atherosclerosis are well known [11–13]. Furthermore, lipids are key factors in cell signalling [14] and control membrane traffic [15]. Hence, the detection of many novel mechanisms by lipidomics studies can be anticipated [16].

Corresponding author. Gerhard G. Thallinger, Institute for Genomics and Bioinformatics, Graz University of Technology, 8010 Graz, Austria. Tel.: +43-316-873-5343; Fax: +43-316-873-105343; E-mail: Gerhard.Thallinger@tugraz.at

Jürgen Hartler is a postdoctoral researcher at the Austrian Centre of Industrial Biotechnology and the Institute for Genomics and Bioinformatics, Graz University of Technology. His research focuses on the development of software applications to foster high-throughput analysis of biomolecules, particularly analyses by mass spectrometry.

Ravi Tharakan is a PhD student at the Johns Hopkins University School of Medicine, Cellular and Molecular Medicine Graduate Training Program. His research interests are neuroimmunology and neuropathology, as well as development of bioinformatics tools to mine ‘omics’ datasets.

Harald C. Köffeler is director of the core facility for mass spectrometry and lipidomics at the Medical University Graz. His research interest is focused on development of high resolution lipidomic LC-MS/MS methods.

David R. Graham is an assistant professor at the Johns Hopkins School of Medicine in Baltimore MD, USA. His research interests include using multomics approaches in the understanding of the role of lipid rafts in HIV pathogenesis and other physiological processes including cardiovascular disease.

Gerhard G. Thallinger is Head of Bioinformatics at the Institute for Genomics and Bioinformatics, Graz University of Technology and director of the Core Facility Bioinformatics at the Austrian Centre of Industrial Biotechnology. His research focuses on the analysis of data from high-throughput technologies with emphasis on mass spectrometry and next generation sequencing.
The detailed investigation of the lipidome in a high-throughput manner has become generally possible by advances in mass spectrometric technologies [17–21], which outperform other technologies particularly in their sensitivity and specificity [22]. Particularly soft ionization techniques like electrospray ionization (ESI) or matrix-assisted laser desorption (MALDI) increased the sensitivity for low abundance species. These techniques ionize the original lipid molecule by attaching adduct ions without cleaving the parent molecule, permitting the transfer of intact molecules to the mass spectrometer. Consequently, single-stage MS alone allows the identification and quantitation of intact lipid species only. This information is not sufficient for comprehending all underlying biological mechanisms, however, because the chemical structure of lipids often leads to lipid species of the same mass but different structures. The functions and effects of these isobaric species are typically substantially different in metabolism, signaling and the development of diseases. For example, Kritchevsky et al. [23] showed that the position of fatty acids (FAs) at the glycerol backbone (sn position) has an impact on the development of atherosclerosis. Additionally, it was found that the FA composition of triacylglycerols (TG) and especially the FA at the sn-2 position plays a key role in nutrition [24]. This hypothesis is sustained by the fact that ‘healthy’ native vegetable oils contain predominantly unsaturated acids at sn-2 [25]. Another example for the importance of FA position is signaling, where it is commonly assumed that for diacylglycerol signaling, the mono-unsaturated or saturated FAs are at the sn-1 position of the glycerol while the poly-unsaturated ones are at sn-2 [26]. Also important is the detection of the double bond position in the FA chain, because, e.g. different linoleic acids have different effects on adipocytes [27].

Therefore, it is important to discriminate isobaric lipid species based on their underlying structure. MS/MS technology, also termed tandem mass spectrometry, is the major strategy for this. Basically one single m/z value is isolated, fragmented by collision-induced dissociation (CID) and the resulting fragment spectrum is recorded. Identification of compounds is achieved by assignment of structure-specific fragments. Unfortunately, the analysis of fragmentation spectra is not as straightforward as it is in proteomics, where the usual low energy CID cleavage sites of peptides are located at the amide bond between the amino acids, resulting in b and y fragments [28, 29]. These fragmentation patterns are highly conserved throughout peptides and proteins and reveal the amino acid sequence of the protein. In contrast to proteins, lipids show a much higher diversity of fragmentation patterns than amino acids. While the peptide bond is the favored cleavage site in proteins, lipids do not have any single-favored cleavage site or ionization polarity due to the huge heterogeneity of their chemical structures and functional groups. The major determinants for the appearance of lipid MS/MS spectra are (i) type of mass spectrometer, (ii) collision energy [30], (iii) charge state of ions, (iv) ionization mode (positive or negative), (v) adduct ions [31] and (vi) the lipid class under investigation. Thus, quite often a specific experimental setup is required to reveal essential structural information. A plethora of studies has been performed for investigating and improving fragmentation processes, and subsequently for the generation of MS/MS interpretation rules [31–62]. Because of the large number of diverse fragmentation patterns, the development of bioinformatics tools for the interpretation of lipidomics MS/MS spectra is a challenging task.

In this article, we systematically review available tools for the analysis of MS/MS data in lipidomics, including software originating from metabolomics which are applicable for lipids. We first give a brief overview of mass spectrometry technologies, to categorize the tools. Details about current experimental approaches using MS in lipidomics can be found in several reviews [63–66]. This review focuses on MS/MS interpretation tools; computational lipidomics MS analysis tools and other available bioinformatics resources are examined in detail in [67, 68] and [69], respectively.

STRUCTURAL ANALYSIS OF LIPIDS BY MASS SPECTROMETRY

The most abundant lipids in mammalian systems are glycerolipids. Glycerolipids contain glycerol, a C3 carbohydrate, as the central chemical structure. One hydroxy group is attached to each of the three backbone carbons and serves as a link for further chemical moieties. The three central glycerol carbons have a stereo numbering and are referred to as sn-1, sn-2 and sn-3 position. The chemical moieties attached to the three hydroxy groups by ester bonds are usually either FAs or a phosphate group (Figure 1). The phosphate group is exclusively
Figure 1: Molecular building blocks of glycerophospholipids are FAs esterified at \( \textit{sn}-1 \) and/or \( \textit{sn}-2 \) positions of the glycerol backbone. The \( \textit{sn}-3 \) position is esterified with a polar head group. The high isobaric diversity of lipids is illustrated by some of the possible underlying structural isomers of PI 38:4 (C\(_{47}\)H\(_{83}\)P\(_{1}\)O\(_{13}\)).

- (A) \( \textit{sn}-1 \) FA 18:0, \( \textit{sn}-2 \) FA 20:4, \( \textit{sn}-3 \) InsP;
- (B) \( \textit{sn}-1 \) FA 20:4, \( \textit{sn}-2 \) FA 18:0, \( \textit{sn}-3 \) InsP;
- (C) \( \textit{sn}-1 \) FA 18:1(11Z), \( \textit{sn}-2 \) FA 20:3, \( \textit{sn}-3 \) InsP;
- (D) \( \textit{sn}-1 \) FA 18:1(9Z), \( \textit{sn}-2 \) FA 20:3, \( \textit{sn}-3 \) InsP;
- (E) \( \textit{sn}-1 \) FA 18:1(9E), \( \textit{sn}-2 \) FA 20:3, \( \textit{sn}-3 \) InsP.
attached on sn-3 and often further linked with a polar metabolite (e.g. choline, ethanolamine, serine, inositol). This whole entity is specific for glycerophospholipids and is called polar head group. Depending on the lipid class, FAs can be attached on any of the sn positions. FAs are the core components of lipids because they serve as molecular energy stores and they are essential for generation of hydrophobic cellular confinement structures. FAs show a high degree of variability according to their carbon and double bond number. In mammals, the most common fatty acyl chain lengths range from 12 to 26 carbons with up to 6 double bonds. Glycerophospholipid structures with a polar head group and either one or two FAs at sn-1/sn-2 position are predominantly responsible for maintenance of cellular membrane structures, whereas glycerolipids without polar head group and one, two or three esterified FAs are usually to be found in the realm of energy storage. Additionally, fatty alkyls can be attached to the glycerol backbone via an ether bond, plasmalogens being a typical example. If a lipid species is identified by single stage MS only (no fragmentation), the result typically contains the lipid class, the number of carbon atoms in the FA chains, and the number of double bonds [70]; e.g. PI 38:4 corresponds to phosphatidyl inositol containing the PI head group at sn-3, 38 carbon atoms in the two chains at sn-1 and sn-2, and 4 double bonds in these 2 chains. The number of carbon atoms in the individual chains remains unknown by single stage MS, as does, not surprisingly, the position of double bonds. Conventional MS/MS can determine the respective lipid class, the individual FA chains, and in many cases their position on the glycerol backbone (Figure 2), if there is sufficient knowledge about the fragmentation process for the experimental setup.

There are essentially two MS approaches in lipidomics (Figure 3), which are preferably used in combination with the soft ionization technique ESI: (i) the shotgun approach that analyzes samples directly from crude extracts and (ii) the

![Figure 2: MS/MS spectrum of PI 18:0/20:4 (m/z 885.5) acquired with an LTQ ion trap mass spectrometer in negative ESI. Ions 1, 2 and 3 correspond to inositol phosphate ions, which are characteristic for the polar head group. Ion 4 represents inositol phosphate with the glycerol backbone attached and is thus also characteristic for PI. The ions labeled with 5 and 6 are the carboxylates of FA 18:0 and FA 20:4, respectively. The ion cluster at 7–10 represents neutral losses of FA 20:4 and FA 18:0, either as FAs (7, 9) or as ketenes (8, 10). Finally, peaks 11 and 12 derive from neutral loss of either FA 20:4 or FA 18:0 followed by an additional loss of inositol from the head group. According to Hsu and Turk [3], two ion pairs are specific for the location of the FAs. (i) Charge driven processes favor the neutral loss of FA + inositol at sn-2 for PI. II > I2 indicates FA 20:4 to be on sn-2. (ii) Secondary fragmentation processes favor the formation of carboxylates of FA at sn-1. 5 > 6 indicates FA 18:0 to be on sn-1. According to this spectrum, the observed lipid is PI 18:0/20:4, and not PI 20:4/18:0.](http://dx.doi.org/10.1093/bib/bbs054)
chromatography approach that pre-separates the analytes before online infusion in the mass spectrometer. For the latter, liquid chromatography (LC) is the preferred technology for lipidomics due to its high sensitivity and its applicability to a variety of lipid classes [71]. Accordingly, only tools for LC–MS analysis will be discussed.

Shotgun mass spectrometry in lipidomics has advantages in its relatively easy experimental setup and rapid quantitative analysis of major lipid species [64]. The easy quantitation is a result of the infusion of the whole extract at once, thus supplying stable signal intensities. However, the simultaneous injection of all analytes causes suppression of low abundance species during the ionization process. Nevertheless, low abundance species can be detected by MS/MS technologies [72] using precursor ion scans (PIS), neutral loss scanning (NLS) and single reaction monitoring (SRM; or multiple reaction monitoring - MRM) (Figure 4). PIS and NLS are full scan methods mainly offered by triple quadrupole or quadrupole time-of-flight (Q-TOF) MS devices. These methods only detect compounds that create defined fragments in MS/MS scans [65], which can be based on polar head groups, fatty acyl moieties or even specific backbone features. In the case of PIS, all precursor masses of a certain fragment ion are detected, whereas in NLS Mass Analyzer 1 and 3 are simultaneously operated at a certain mass offset equivalent to the uncharged molecular weight of the neutral loss under investigation. In contrast, SRM operates at fixed \( m/z \) values for Analyzer 1 and 3, resulting in highly specific precursor/fragment pairs, which is in many cases indicative for just one compound. The drawback of SRM is its targeted nature, which does not allow detection of unanticipated lipid species. Examples for combination of different PIS/NLS experiments are the multiple precursor ion scanning (MPIS) approach developed in the Shevchenko lab [73, 74] and the multi-dimensional mass spectrometry-based shotgun lipidomics method developed by Han et al. [65]. Product ion spectra on each lipid compound of interest are usually the best way to fully characterize a precursor ion [58, 75].

LC–MS in lipidomics is characterized by at least one additional layer of separation preceding \( m/z \) analysis. A chromatographic separation step substantially increases the number of detectable lipids due to reduced suppression effects in the ion source [76–78]. In this manner, e.g. for reverse phase LC–MS, the identification of very low abundance analytes is possible without any manual intervention in the analysis process [79], which is especially useful for high-throughput platforms. An impressive example for the sensitivity of the LC–MS approach is the detection of a minor class of ether-linked TG by normal-phase chromatography followed reversed phase LC–MS [80]. In addition to \( m/z \) values,
LC–MS also offers retention time values for identification purposes. Targeted SRM can be used with LC–MS as well, as shown by the LIPID Metabolites And Pathways Strategy (LIPID MAPS) consortium [81, 82]. A successful example for high resolution full scan LC–MS is shown by Koulman et al. [83]. Although their instrumentation lacked MS/MS capabilities, sub-ppm mass accuracies and retention time were sufficient for identification of certain glycerophospholipid classes in fast survey scans. Higher confidence in identifications can be achieved when MS/MS spectra are additionally acquired and used during data analysis. On LTQ-Orbitrap and LTQ-FT instruments, simultaneous acquisition of high resolution precursor spectra and low resolution MS/MS product spectra is possible, resulting in three layers of information for any given compound: exact mass, retention time and structural features [84–86].

Irrespective of the MS approach (shotgun or LC), structural information on the lipid species can only be revealed by information on the fragmentation process of the lipid classes of interest. However, the appearance of an MS/MS spectrum can vary tremendously depending on the mass spectrometer used, the collision energy [30], charge state of analyte, ion mode (positive or negative) and the adduct ions used [31]. Hsu, Turk and co-workers [31–36, 38–40, 43, 44, 46, 48–51, 53–62] have published many fragmentation rules readily applicable for automated analysis tools. However, these rules are not sufficient for detecting the positions of the double bonds at the FA chains. Two novel strategies tackle this challenge. The first one, Ozone–Induced–Dissociation (OzID), creates characteristic fragmentations revealing the double bond-positions [87–90]. It is based on the fact that ozone cleaves the double bonds in ionized lipids, whereupon the fragments are detectable in the MS scans. The second strategy relies on the principle that fragments in MS/MS reveal information about the double bond position via charge remote fragmentations [57, 91–95], whereupon either high energy CID or spectra [C21]MS3 are required.

**MS/MS ANALYSIS TOOLS FOR LIPIDOMICS**

The way available bioinformatics tools tackle complex MS/MS analysis for lipidomics is manifold. Existing solutions are mostly tailored to the specific requirements of a specific laboratory; therefore, generally applicable software is scarce. Thus, to establish a largely automated high-throughput platform to obtain structural information about lipids, the experimental strategy strongly depends on the mass spectrometer used, the lipid classes of interest, the MS/MS analysis strategy and also the analysis tool. Despite the variety of available applications, the underlying principles can be used for categorization into four groups: (i) tools using reference libraries and spectrum similarity searches; (ii) tools for the analysis of shotgun data; (iii) tools for the analysis of LC–MS data and (iv) tools providing simple MS/MS searches, where the focus is to give hints toward possible structures and annotations of candidate hits. In this section, these categories and their available tools are examined in detail; an overview is given in Table 1.
Table I: Overview of available MS/MS analysis tools.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Ref. spec.</th>
<th>Customizable</th>
<th>Availability</th>
<th>MS</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMDMS-SL</td>
<td>Shotgun</td>
<td>No</td>
<td>Yes</td>
<td>Free</td>
<td>High, low</td>
<td><a href="http://shotgunlipidomics.com/programs/programs.htm">http://shotgunlipidomics.com/programs/programs.htm</a></td>
</tr>
<tr>
<td>HMDM</td>
<td>Similarity</td>
<td>Yes</td>
<td>No</td>
<td>Web-service</td>
<td>High, low</td>
<td><a href="http://www.hmdb.ca">http://www.hmdb.ca</a></td>
</tr>
<tr>
<td>LipidInspector</td>
<td>Shotgun</td>
<td>No</td>
<td>No</td>
<td>Upon request</td>
<td>Low</td>
<td><a href="http://msr.dom.wustl.edu/Personnel/StaffScientistSongHaowei.htm">http://msr.dom.wustl.edu/Personnel/StaffScientistSongHaowei.htm</a></td>
</tr>
<tr>
<td>LipidQA</td>
<td>Shotgun</td>
<td>Yes</td>
<td>No</td>
<td>Free</td>
<td>Low</td>
<td><a href="http://metabolism.scripps.edu/">http://metabolism.scripps.edu/</a></td>
</tr>
<tr>
<td>Lipid Search</td>
<td>LC</td>
<td>Adjust</td>
<td>No</td>
<td>Web-service</td>
<td>High</td>
<td><a href="http://metabolism.scripps.edu/">http://metabolism.scripps.edu/</a></td>
</tr>
<tr>
<td>Lipid View</td>
<td>Shotgun</td>
<td>No</td>
<td>No</td>
<td>Commercial</td>
<td>Low</td>
<td><a href="http://www.absciex.com/products/software/lipidview-software/">http://www.absciex.com/products/software/lipidview-software/</a></td>
</tr>
<tr>
<td>LipidXplorer</td>
<td>Shotgun</td>
<td>No</td>
<td>Yes</td>
<td>Open-source</td>
<td>High, low</td>
<td><a href="http://sourceforge.net/projects/lipidxplorer">http://sourceforge.net/projects/lipidxplorer</a></td>
</tr>
<tr>
<td>MassBank</td>
<td>Similarity</td>
<td>Yes</td>
<td>Yes</td>
<td>Open-source</td>
<td>Low</td>
<td><a href="http://www.massbank.jp/">http://www.massbank.jp/</a></td>
</tr>
<tr>
<td>Profiler-Merger-Viewer</td>
<td>LC</td>
<td>No</td>
<td>Yes</td>
<td>Upon request</td>
<td>High</td>
<td><a href="http://metabolism.scripps.edu/">http://metabolism.scripps.edu/</a></td>
</tr>
<tr>
<td>XCMS2</td>
<td>Similarity</td>
<td>Yes</td>
<td>Yes</td>
<td>Free</td>
<td>High, low</td>
<td><a href="http://metabolism.scripps.edu/xcms/">http://metabolism.scripps.edu/xcms/</a></td>
</tr>
<tr>
<td>Cyberlipid Center</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://www.cyberlipid.org">http://www.cyberlipid.org</a></td>
</tr>
<tr>
<td>GOLM</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://gmd.mpimp-golm.mpg.de/">http://gmd.mpimp-golm.mpg.de/</a></td>
</tr>
<tr>
<td>LIPIDAG</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://www.lipidag.tcd.ie/">http://www.lipidag.tcd.ie/</a></td>
</tr>
<tr>
<td>LIPIDAT</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://www.lipidat.tcd.ie/">http://www.lipidat.tcd.ie/</a></td>
</tr>
<tr>
<td>Lipid Bank</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://lipidbank.jp/">http://lipidbank.jp/</a></td>
</tr>
<tr>
<td>Lipid Library</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://lipidlibrary.aocs.org/">http://lipidlibrary.aocs.org/</a></td>
</tr>
<tr>
<td>LMSD</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>N/A</td>
<td><a href="http://www.lmsd.tcd.ie/">http://www.lmsd.tcd.ie/</a></td>
</tr>
<tr>
<td>LIPID MAPS</td>
<td>Database</td>
<td>No search</td>
<td>No</td>
<td>Web-service</td>
<td>High, low</td>
<td><a href="http://www.lipidmaps.org/">http://www.lipidmaps.org/</a></td>
</tr>
</tbody>
</table>

The column ’Ref. spec.’ indicates whether the MS/MS search is based on experimental reference spectra, whereupon ‘adjust’ means that experimental reference spectra were used for adjusting theoretical ones. ’Customizable’ stands for customization options for self-defined lipid species and potential extensibility. MS reflects the resolution type of the mass spectrometers to which the software is dedicated. High encompasses MS instruments such as FT and Orbitrap, and low encompasses low-to-medium resolution that is provided by instruments such as ion trap, triple quadrupoles and Q-TOF.

Spectral similarity search tools

Common to all the spectral similarity search tools are the following steps: (i) compilation of a database containing reference spectra; (ii) filtering of these spectra by precursor $m/z$; (iii) similarity search between experimental spectrum and remaining reference spectra and (iv) scoring and output of potential identification candidates. This strategy is quite similar to proteomics search engines [96–100] except that the spectral reference database in proteomics is in silico calculated from protein sequence databases, rather than determined empirically. Even though the tools presented here have been primarily developed for metabolomics, they contain a certain amount of lipid spectra. Thus, searches of uncharacterized spectra versus these databases can give useful hints regarding potential candidate lipid species.

The first tool, XCMS2 [101], features the METLIN database [102], which contains a library of Q-TOF MS/MS spectra. The XCMS2 similarity search relies on a modified version of the ‘shared peak count’ method that takes precursor mass filtering into account. The algorithm generates a similarity matrix and a distance matrix between the observed and the reference spectrum and calculates corresponding scores. Of these two scores, a percentage match score is calculated that defines whether the reference component is a potential match. The apparent restriction to Q-TOF data limits the applicability of XCMS. An array of metabolomics experiments is publically available through XCMS2 online at (https://xcmsonline.scripps.edu/).

The second tool is the Human Metabolome Database (HMDB) [103, 104]. HMDB offers more MS/MS spectra of lipid species than XCMS2 at this time. Furthermore, HMDB includes experimental MS/MS spectra for Fourier Transform–Ion Cyclotron Resonance–MS (FT–ICR–MS), Ion Trap, Q-TOF and Triple Quadrupole instruments at various collision energies. Interestingly, HMDB supports searches for NMR spectra as well. According to Wishart et al. [103], the scoring strategy is similar to a previously published proteomics scoring scheme [105], whereas details about the metabolomics search are not published. In that article, the
similarity was calculated by a linear equation from weighted values of SEQUEST [96] result parameters. The weighting coefficients were determined by multivariate discriminant analysis of a training data set of more than 3000 MS/MS spectra. General applicability of HMDB is reduced by its restriction to human metabolites only.

The third tool is the MassBank [106]. It contains both MS and MS/MS spectra of various metabolites including phospholipids. The similarity search relies on a modified version of the cosine correlation algorithm (described by Stein et al. [107]). However, the best reported positive predictive value was ~50% [106], which limits MassBank’s applicability in high-throughput data analyses. Nevertheless, MassBank can be a useful instrument to detect candidate species or candidate fragments. In contrast to other databases, MassBank allows the contribution of self-created reference spectra.

The MS/MS tools presented so far are able to identify the lipid FA chains by similarity search against a database of precompiled spectra. Generally, the use of a spectral database has the drawback that the acquired spectra are often limited to specific types of mass spectrometers and experimental setup parameters such as collision energy and adduct ions used, as the appearance of fragmentation spectra can vary tremendously depending on these parameters. For example Hsu and Turk [31] demonstrated that an ITMS instrument returns different ions containing structural information than a TSQ instrument, and that adduct ions have a major influence. Furthermore, the use of reference libraries is especially questionable in shotgun experiments, because mixed spectra containing more than one isobaric precursor occur frequently [108]. Moreover, such approaches can detect only lipids that are already part of the mass spectral database. The structural elucidation of newly identified components is not always possible, because the databases can hardly be extended by end-users. In this respect, XCMS2 and MassBank have the advantage of being extensible by self-created reference spectra while HMDB provides just the download of the spectral database and not the tool itself. However, in any case, the quality of the detection relies on the existence of reference spectra, and it is hard to foresee which components are required for unknown biological samples. Thus, these solutions are of limited use in experiments where the lipid species are unknown a priori.

Shotgun MS/MS tools
In shotgun analyses, the whole sample extract enters the mass spectrometer without prior separation of compounds. The specificity and sensitivity of this technique derives from the application of MS/MS, where PIS, NLS and SRM are preferably used. A usual strategy is to scan for class-specific head groups and probable FA chains by PIS or NLS. Correspondingly, in this group, tools dealing with structural elucidation were reported relatively early. Moreover, for many lipid classes, these tools are even able to determine the position of the FAs at the glycerol backbone, i.e. the regio-specific position, by comparison of fragment intensities. Generally, for this category many more MS/MS high-throughput analysis tools are available than for the others. The reason for this is presumably the appealing analytical simplicity of the method and the fact that it derives its analytical power from MS/MS.

The first representative in this category is the LipidProfiler software package [73]. This tool exploits the MPIS approach, where several PIS of multiple precursors are generated in parallel. By scanning for head groups and FAs, the detection of many co-existing isobaric substances is possible. Regiospecificity is defined for many analytes as well. The package was developed in collaboration with AB Sciex, which limits its applicability to vendor-specific MS instruments. Additionally, it is not extensible with newly identified lipid species, although the software offers 44 lipid classes containing 25 000 lipid species [109]. Currently, the software is marketed under the name LipidView.

A drawback of the MPIS technology is that simultaneous parallel acquisition of NLS is not possible, which is necessary for detection of the head groups of phosphatidyl ethanolamine and phosphatidyl serine. To circumvent this problem, an MS strategy was developed where a survey scan of 2 s is employed, followed by two consecutive MS/MS scans in data-dependent acquisition (DDA-fragment spectra are automatically acquired by the instrument from the most abundant molecular ions determined in a preceding single stage MS spectrum) mode to cover PIS and NLS [74]. This kind of data can be analyzed with LipidInspector [74], which was developed by Scionics Computer Innovations and is available upon request from the authors.

The LipidQA software platform [110] is dedicated to the analysis of MS/MS spectra of product ion scans in DDA mode, which differs from the
traditional shotgun approach of the previous two tools. The package was tested on low-to-medium resolution instruments (linear ion trap, triple quadrupole, Q-TOF). The strategy is similar to the spectral similarity search tools previously described, because it also relies on a library of reference spectra. The similarity search is based on the so-called ID score which is a measure of the percentage of fragment ion m/z values of the 40 most intense peaks matching to the theoretical spectrum. Furthermore, the software detects the regio-specific sn-positions. The development team of LipidQA is itself engaged in investigations of novel lipid fragmentation pathways [31–36, 38–40, 43, 44, 46, 48–51, 53–62, 93], which is certainly beneficial for the quality of reference spectra. However, the software is currently available for phospholipids and sphingomyelins only.

The fourth tool pursues the strategy of multidimensional mass spectrometry (MDMS), where the analysis is based on a series of MS, PIS and NLS scans that are acquired under varying experimental conditions [111] (a review about this technique has been published recently [65]). The AMDMS-SL package [112] exploits the information in a two-step procedure, which quantitates unambiguously detectable high abundance components first, and then uses these results to quantitate low abundance or overlapping species. The identification is based on an extensible ‘building block’ library containing information about backbone, head group, adduct ions, ionization mode and a formula for the FA chains. This formula incorporates the elements for the possible connection to the backbone and variables for the carbon atoms and the number of the double bonds in the chain (ranges for variables are customizable). Out of this information, theoretical lipid species are calculated and used for the search. Furthermore, class-specific fragments and losses can be defined for PIS and NLS, respectively. Mass spectrometrically, a lipid class specific ionization is required. The acquired MS, PIS and NLS scans are searched against the theoretical lipids, which results in a candidate peak list for each lipid class. Then, to identify the FAs, theoretical fragments are searched against PIS and NLS spectra, where a comparison of the intensity of the fragments is used to determine the regio-specific position. The proposed concept is attractive due to its extensibility and can be seen as the first generic attempt for the structural elucidation of lipid species.

The LipidXplorer package takes a further step into this direction by introducing the concept of the molecular fragmentation query language (MFQL) [108]. The control for lipid species identification is transferred completely to the user. The software organizes MS and MS/MS spectra of all acquired experiments in database-like structure in a flat-file format. MFQL queries are used to generate m/z values of signal ions to search this database. The queries are defined for each lipid class and can contain similar ‘building blocks’ such as in AMDMS-SL (head group, FA chain constraints, etc.), but they are not exclusively restricted to these. For example, AMDMS-SL looks for certain precursor masses in MS for high-resolution data and checks for expected chain fragments in PIS and NLS to assign the FA chains. In LipidXplorer, the user can define in the query which check is to be performed at which MS level. Furthermore, these rules can be combined by Boolean operators. In this manner, the algorithm is no longer restricted to certain signal ions for determination of the FA chains. This concept ensures independence of MS/MS scan types (such as PIS, NLS, SRM and DDA), MS instrumentation and experimental setup parameters such as collision energy and adduct ions [113].

In summary, in shotgun-based lipidomics, excellent tools are available for the analysis of MS/MS spectra. Shotgun tools can identify the FA composition even for isomeric structures, where in many cases the regio-specific position can be detected too. Furthermore, some approaches are more generally applicable already, showing a trend leading away from applications tailored to specific devices and experimental setups. The major disadvantage lies in the shotgun approach itself, because mixed spectra of isomeric and isobaric substances occur frequently without chromatographic separation. However, this problem is tackled by smart application of PIS, NLS and SRM, in many cases.

**LC–MS/MS tools**

In contrast to shotgun lipidomics, in LC–MS the sample is released gradually into the mass spectrometer. The advantage is an additional layer of separation that even permits the segregation of positional isomers [114]. Additionally, the analytical distribution of the analytes allows the ion source to supply more ions, and thereby reduces the suppression of low abundance species. Furthermore, the MS/MS spectra in LC–MS are usually product ion scans,
which contain all the fragment ions yielded by the collision process. Because the analytes enter the mass spectrometer gradually, the time for detecting MS/MS spectra is limited. Consequently, only the most intense peaks are selected for MS/MS analysis, whereas structural information of low abundance species is absent. Moreover, software tools for LC–
MS/MS analysis of lipids are scarce.

The Profiler–Merger–Viewer tool provides a semi-automated analysis that requires a specialized experimental setup, and is applicable to high-resolution mass spectrometers only [115]. The mass spectrometer performs a full scan followed by two single ion-monitoring (SIM) scans of the two most abundant peaks. From these peaks MS² and MS³ scans are conducted. The software searches the SIM scans for potential hits. Then, the predictable existence of possible neutral losses in MS² is used as an additional exclusion criterion for false positives. The MS² and MS³ scans are used to determine the potential FA composition, whereupon the regio-specific position is not reported. The output is a list of chromatograms of potential candidates that must be manually selected for quantitation. The software is easily extensible, but manual peak selection can be cumbersome for datasets from high-throughput studies. Furthermore, the restriction to high-resolution data and the specialized experimental setup limits its field of application considerably.

The Lipid Search tool [85] contains a database of more than 200 000 theoretical m/z values for lipids and their corresponding theoretical fragments. Typically, the m/z values for the fragments are created in silico, and they are additionally adjusted by experimentally obtained data. The MS/MS detection includes parameters such as scan type, adduct ions and m/z tolerance; however, details about the search algorithm are not reported. This approach proved its value in the detection of various lipid classes of phospholipids from tissue samples [85]. Remarkably, even positional isomers and low abundance alkylacyl and alkenylacyl isobaric species were detected. However, Lipid Search was primarily used as identification tool in this study, while quantities were obtained by manual calculation. The tool is provided as a web application and currently covers phospholipids only. No means for extension to other lipid classes are available.

LC–MS/MS bears the potential for detecting very low abundance lipid species due to the chromatographic separation and concentration of the lipids. However, available tools are not broadly applicable because of their limitation to specific lipid classes and MS devices, and their semi-automated nature. Generic solutions as available for the shotgun approach are desirable for this technique. MS/MS extensions to tools that already achieve good sensitivity and specificity by MS analysis [116, 117] give rise to novel analytical perspectives, particularly because LC–MS/MS is more suited to top–down research than the shotgun technique.

**Annotation databases for lipids**

After identification of lipid compounds, the species must be annotated with respect to structure and function. Several tools are available for the annotation of lipids identified by LC– or GC–MS. The LIPID MAPS database provides the broadest range of tools [118]. The service is intended to map the lipidome in a systems biology based manner, comprising a comprehensive ontology and classification system for lipids [119, 120]. There, lipids are stratified into eight classes: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, polyketides and the neologism saccharolipids, which includes all FAs linked to a sugar backbone. This new term was introduced because the term ‘glycolipids’ is ambiguous, as all lipid classes include glycan derivatives. Additionally, a standardized set of guidelines for drawing structures of lipids is available.

For the purpose of further characterization of identified lipid species, LIPID MAPS introduces a comprehensive suite of annotation tools. First, a basic search function is provided, which allows a user to search a compound of interest against several different databases: a lipid database search (LMSD) [121] for direct lipid annotation, a molecular structure-based search and a lipid standards search. Useful for laboratories with bioinformatic personnel, all the database searches can be accessed programmatically. Finally, LIPID MAPS also contains tools for searching mass spectrometry results, an outline of the lipid classification scheme, a database of published lipidomics datasets according to biology, tools for creating pathway diagrams from lipid datasets and structure drawing tools.

A second database that may be useful for lipid annotation is the GOLM metabolome database [122, 123]. This database is designed for GC–MS studies of metabolite compounds, and uses mass spectral tags, containing spectra and retention time, to store compound identifications, as well as a data
model for chemical substances. The database is primarily a metabolite database, and so may be less appropriate for lipid searches. Good annotations are available for lipids, including stereo-isomer and isotope-momer information, molecular and monoisotopic masses, along with link-outs to KEGG [124] and ChemSpider [125], both of containing a large amount of information on biochemical pathway involvement and chemical structure, respectively. An interesting feature of this database is the ability to use decision trees to predict the functional groups of compounds. However, this latter service does not seem to be fully operational yet.

Another database useful for functional annotation of lipid species is LipidBank [126], the database of the Japanese Conference on the Biochemistry of Lipids. The database provides a search engine, similar to LIPID MAPS, which takes as input a LipidBank ID, a lipid name, or formulae and a molecular weight range. Furthermore, one interesting feature is that the database is also searchable by chemical synthesis, source or biological activity. This gives an easy way to mine pathways directly through the identifications database. The database also provides archived spectral data, such as UV spectra, IR spectra, NMR spectra and MS spectra, along with chromatogram data and some physical properties.

Finally, there are several other lipid knowledge bases, which may be useful for manual annotation of lipid identifications. These include The Lipid Library [127], Cyberlipid Center [128] and the Caffrey Labs collection of databases [129], which include LIPIDAT [130], a database of thermodynamic information on many lipid species, LIPIDAG [131], a database of lipid miscibility, LMSD [132], a database of lipid structures and CMCD, a database of critical micellar concentrations. Each of these databases may provide useful information not only for lipid annotation but also for basic lipid analysis techniques and chemical biology. However, several of these databases are no longer being updated, and much of the relevant information from them seems to have already been integrated into more current databases, such as LIPID MAPS.

CONCLUSIONS

Lipidomics is a young discipline among the ‘omics’ sciences, but it has been recognized as an essential complement to well-established disciplines such as transcriptomics and proteomics. Although mass spectrometry has been a crucial factor for high-throughput lipidomics, the elucidation of lipid structure is far from trivial. Detailed structural assignment requires sophisticated MS/MS approaches. However, available analysis software is often tailor made and consequently only useful for a specific experimental setup. Exceptions exist in shotgun analysis [108, 112, 113] where special emphasis was placed on general applicability and easy extensibility for end-users. Similar applications are not yet available for LC–MS/MS, where solutions are limited to specific lipid classes or experimental setups, and fully automated analysis is not supported. Compounds not covered by the tools must be interpreted manually or searched by similarity search tools or annotation databases, which primarily allow the analysis of a single spectrum only.

Lipidomics analysis software can determine the FAs composition and regio-specific positions. Standard MS/MS methods do not allow the identification of the double bond positions, but two novel approaches make this possible. The OzID technology [87–90] cleaves C–C double bonds by ozonolysis in the ESI ion source. This leads to characteristic fragments revealing the double bond position in MS scans. However, this method has some drawbacks that complicate bioinformatics analysis: (i) the additional fragments can originate from different intact lipid precursors; (ii) the additional fragments can result in additional overlapping isomers or isobars that have to be classified by MS/MS; (iii) lipids cannot be quantified easily, because the cleaved fragments reduce the intensity of the intact lipid species; the intensity of the fragments, which can contain isomers, has to be considered. To tackle these challenges, MS algorithms will require a much higher level of sophistication that must work in concert with MS/MS fragment detections. The second approach relies on the sophisticated application of MS/MS strategies to detect characteristic fragments for double bonds in MS/MS [57, 91–95]. Computationally, detecting double bond fragments is similar to detecting FA fragments. Thus, presented generic approaches [108, 112, 113] could possibly be of use.

The analytical progress in lipid structural elucidation opens up novel challenges for bioinformatics. Additionally, lipidomics MS and MS/MS trend towards high-throughput analysis [133]. Novel tools or extensions to existing software are required to keep pace with these advancements. For the development...
of such tools, extensibility and general applicability shall in particular be considered in the design, to open them to a wider community. Generic approaches can be more easily adapted to analytical requirements and subsequently, the advantages of the software more quickly conveyed to end-users. Consequently, high-throughput platforms are more rapidly implemented, effectively promoting lipidomics research.

**Key Points**

- **MS** is the analytical method of choice in lipidomics due to its high sensitivity, specificity and large dynamic range, but due to isobaric and isomeric lipid species, **MS/MS** has to be applied for structural elucidation.
- **MS/MS** fragmentation processes of lipids depend on various experimental parameters; consequently, generic bioinformatics solutions are required, to keep pace with analytical advancements.
- Available software tools can be classified into four categories: (i) tools using reference libraries and spectrum similarity searches; (ii) tools for the analysis of shotgun data; (iii) tools for the analysis of **LC-MS** data; (iv) tools providing simple **MS/MS** searches, where the focus is structure annotation.
- Some generic bioinformatics solutions for shotgun lipidomics exist, whereas the **LC-MS** approach is poorly supported.
- Extensibility and general applicability must be a key consideration in the conception of novel **MS/MS** tools to keep pace with analytical advances.

**Acknowledgements**

The authors thank Gerald N. Rechberger and Fritz Spener from the Lipidomics Research Center Graz for helpful discussions and encouragement.

**FUNDING**

This work was supported, in part, by the Austrian Ministry of Science and Research GEN-AU project BIN (FFG Grant 820962) and by the EU-FP7 project LipidomicNet (EC Grant 202272). The Austrian Centre of Industrial Biotechnology (ACIB) contribution was supported by FFG, bmvitt, mvwfi, ZIT, Zukunftsförderung Tirol and Land Steiermark within the Austrian COMET program (FFG Grant 824186).

**References**

25. Karupiah T, Sundram K. Effects of stereospecific positioning of fatty acids in triacylglycerol structures in native and
54. Hsu FF, Turk J, Williams TD, et al. Electrospray ionization multiple stage quadrupole ion-trap and tandem quadrupole mass spectrometric studies on phosphatidylglycerol from...


87. Thomas MC, Mitchell TW, Blanksby SJ. Ozonolysis of phospholipid double bonds during electrospray ionization:


118. LIPID MAPS. http://www.lipidmaps.org (22 February 2012, date last accessed).


130. LIPIDAT. http://www.lipidat.tcd.ie (22 February 2012, date last accessed).

131. LIPIDAG. http://www.lipidag.tcd.ie (22 February 2012, date last accessed).

132. LMSD. http://www.lmsd.tcd.ie (22 February 2012, date last accessed).