Lipid identification using a MS/MS database of 120,000 tandem mass spectra

Tobias Kind, Kwang-Hyeon Liu, Do Yup Lee and Oliver Fiehn

UC Davis Genome Center - Metabolomics, Davis, CA

Introduction

Lipid profiling or shotgun lipidomics of polar lipids can be performed using liquid chromatography coupled to mass spectrometry or direct infusion mass spectrometry. In an approach similar to peptide sequencing, precursor information and data-dependent MS/MS scans from iontrap, Orbitrap or hybrid time-of-flight analyzers can be used for mass spectral library search. However the identification process is hampered due to the absence of a universal tandem mass spectral library for lipids. Even large databases such as the NIST08 and Massbank only contain 14,802 and 8,337 tandem mass spectra respectively and around 100 lipid related MS/MS spectra. We introduce an in-silico generated database of 120,000 tandem mass spectra that was developed using high-resolution and low resolution mass analyzers and was validated on literature and in-house data.

Methods

A lipid structure database was developed using combinatorial synthesis programs and compound structures from the LipidMaps database. Mass spectral fragmentation was modeled according to fragmentation observed from both in-house MS/MS spectra and literature data. Several polar lipid classes were included, such as glycerophosphocholines (PC), glycerophosphoethanolamines (PE), glycerophosphoserines (PS), glycerophosphoglycerols (PG), glycerophosphoinositols (PI), sphingomyelins (SM), ceramides (CE), galactolipids as well as diacylglycerols (DAG) and triacylglycerols (TAG). For each of those lipid classes positive and negative MS/MS libraries with different $[M+H]^+$, $[M+Na]^+$, $[M+NH4]^+$ or $[M-H]^-$ electrospray adducts were generated. Fragment peak abundance data was modeled using heuristic rules. Individual lipids are identified using accurate precursor ion filtering and subsequent tandem mass spectral library search. The identification process was validated using reference standards as well as decoy database search strategies

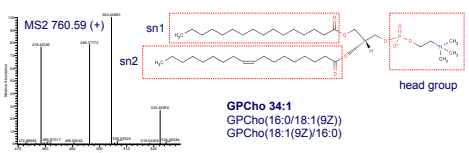
Results

All peaks are stored with accurate mass data and are fully annotated with a LipidMaps nomenclature name, molecular formula, fatty acid and fragment loss explanation. Each in-silico spectrum has an assigned accurate mass precursor ion that is used for precursor filtering during the database matching process. A MS/MS precursor search tolerance of 0.4 m/z units is sufficient for low resolution and high resolution data. After the precursor filter is passed, each spectrum is searched using a dot-product and probability search based method. A library hit score is obtained for all candidates in an approach analogous to mass spectral library search. The database software and graphical user interface is based on the widely used and freely available NIST MS Search GUI program developed by the Chemical Reference Data Group at NIST. The workflow capable command line interface has a library search speed of around 600 spectra per second and can be used for automated searching in large mass spectral datasets.

Platform – nanoESI infusion with iontrap

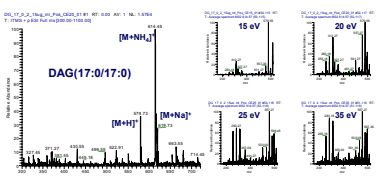
Low-resolution and high-resolution data were obtained from a linear ion trap and a linear iontrap coupled to an Fourier transform (FT-ICR-MS) mass spectrometer. Infusion was performed with a chip-based nanoelectrospray robot (Advion Nanomate). For each injection a new nozzle is used to avoid cross-contamination.

Fatty acid and head group analysis



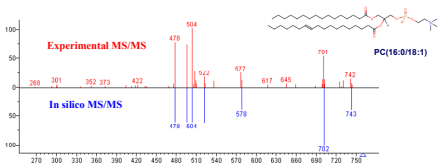
Zoom into the MS/MS data region at 35 eV CID. The four peaks refer to the loss of the sn1 alkyl or acyl group and sn2 group loss. The peaks with delta -18 Da are due to loss of water. The MS/MS analysis additionally can be used to check for correct assignment of possible adducts which would result in different fragmentation patterns. A further assignment of the sn1 and sn2 group would require a MS3 step. In many cases the data dependent MS2 scans are not clean due to selected isolation width of 2 Da and overlap from different lipid species. In this case a quality factor is assigned (low/high).

MS/MS fragmentation voltage experiments



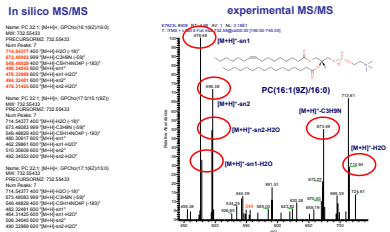
Different lipid species require different MS/MS voltages for abundant fragment generation. The example of the synthetic diacylglycerol DAG(17:0/17:0) standard shows an increased MS/MS fragmentation with increased normalized MS/MS collision energy. For all lipid classes experiments were performed to obtain optimal fragmentation values.

In-silico vs. experimental MS/MS spectra



Reference standard Phosphatidylcholine PC(16:0/18:1) at m/z=760.64 representing $[M+H]^+$. The mass accuracy obtained with the unit mass resolution LTQ iontrap for this example is 36 ppm or 0.027 Da.

Accurate mass MS/MS spectra for matching



The lipid library contains names with stereoinformation, accurate masses for precursor and tandem spectra and data on fragmentations and losses annotated with names and elemental compositions. Additional information can be annotated from MS^3 spectra.

MS/MS search with NIST MS Search GUI

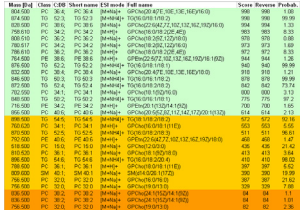
Traditional MS library search allows the deeper inspection and comparison of tandem mass spectra. This approach is superior compared to a manual inspection with simple fragment tables only or black box algorithms. Unknown lipids can be moved to a user library for later inspection. The freely available NIST MS Search program GUI can be used for visual inspection of the library hits. It also can be integrated into existing LC-MS software.

Precursor and product ion m/z tolerances can be adjusted according to high resolution (0.1 m/z unit window) or low resolution instruments (0.4-0.8 m/z unit window) used. After the precursor filter is applied, a dot product hit score and a probability match factor (PBM) and is calculated for each spectrum.

Lipid species covered in database

Lipid class	Short	Species	Spectra	Positive mode	Negative mode
Glycerophosphocholines	PC	5,476	10,952	M+H, M+Na	-
Glycerophosphoethanolamines	PE	5,476	16,428	M+H, M+Na	M-H
Glycerophosphoserines	PS	5,123	10,246	M+H	M-H
Glycerophosphoglycerols	PG	5,476	5,476	low abundant	M-H
Glycerophosphoinositols	PI	5,476	5,476	low abundant	M-H
Glycerophosphates	PA	5,476	16,428	M+NH4, M+Na, M+2Na	M-H
Sphingomyelins	SM	168	336	M+H, M+Na	-
Diacylglycerols	DAG	1,764	-	M+NH4, M+H, M+Na	-
Triacylglycerols	TAG	2,640	5,280	M+NH4, M+H, M+Na	-
Monogalactosyldiacylglycerols	MGDG	5,476	21,904	M+Na	M-H
Digalactosyldiacylglycerols	DGDG	5,476	10,952	M+Na	M-H
Sulfoquinovosyldiacylglycerols	SQDG	5,476	5,476	-	M-H
Hexaacyl Lipid-A (PP)	LipidA-PP	15,625	15,625	-	M-H
SUM		**69,128**	**124,579**		

Library MS/MS search result hit list



Example output for automated NIST MS library search of a plasma sample using accurate precursor identification and MS/MS dot product library matching. The library hit scores above 600 are colored green, orange hits cores from 300-600 (need further inspection) and scores below 100 are possibly wrong identifications.

Conclusions

Around 150 lipid compounds can be annotated using low-resolution iontrap data obtained from human blood plasma.

Whereas the use of triple-quadrupole mass spectrometers for lipid profiling routinely allows the annotation of a higher number of compounds, our new MS/MS library is especially useful for commonly and widely used low resolution mass spectrometers.

The ability of adjusting the precursor and product ion search tolerances, together with the dot product search algorithm makes this database useful for all types of mass spectrometers that can generate data dependent MS/MS scans including LC-MS/MS acquisitions.

Lipid class, carbon and double bond number can be identified. Regiospecificity and stereochemistry can not be assigned.

This project was supported by R01 ES013932 NIH/NIEHS.