Lipid identification by tandem mass spectrometry using an in-silico generated fragment database

Tobias Kind[1], Kwang-Hyeon Liu[1,2], Do Yup Lee[1], Oliver Fiehn[1]

[1] UC Davis Genome Center - Metabolomics, Davis, USA
[2] Inje University - PharmacoGenomics Research Center, Busan, South Korea

Introduction

Lipid profiling or shotgun lipidomics of glycerophospholipids can be performed using direct infusion mass spectrometry and liquid chromatography coupled to mass spectrometers using electrospray ionization. The use of iontrap, hybrid time-of-flight and Orbitrap analyzers for lipid analysis was limited due to the absence of a large and universal mass spectral library of lipids. Using tandem mass spectrometry it is possible to identify major phospholipid head groups together with the correct acyl and alkyl side chain modifications and their double bond position and degree of unsaturation. We present an in-silico tandem mass spectral library of 70,000 lipid mass spectra for glycerolipids generated from LipidMaps structures. The library covers 37,000 lipid species including glycerophosphocholines, glycerophosphoserines, glycerophosphoglycerols, glycerophosphoinositols, glycerophosphoethanolamines, glycerophosphates as well as sphingomyelins and triacylglycerides.

Experimental setup

Lipid Standards for multiple lipid classes were obtained from Avanti Polar and directly dissolved in solvent. Plasma lipid were extracted using methyl-tert-butyl ether (Matyash et al., J. Lip. Res. 2008). In brief, methanol (225 µL) was added to blood plasma, and then 750 µL of MTBE was added and the mixture was shaken. Phase separation was induced by adding 187.5 µL of water and later the sample was vortexed. After shaking, the sample was centrifuged at 14,000g for 2 min. The upper organic phase was collected and dried in a vacuum centrifuge. After adding 10 µL of 100 mM ammonium acetate to 90 µL of the supernatant the lipid extracts were infused into a linear ion trap mass spectrometer using a Advion Nanomate chip-based infusion system (nanoESI). Mass spectra were acquired from 400 to 1,000 Da (+/-) at 0V CID voltage. Data dependent MS2 scans were performed on the 50 highest MS1 peaks and scanned in low-resolution mode with 15V, 20V, 25V and 35V CID voltage to obtain specific MS/MS fragmentations.

Computational methods

Lipid structures were generated using LipidMaps Tools. Around 400,000 chemical structures were stored in SDF files and imported into the freely ChemAxon Instant-JChem structure database. The dataset containing meta data descriptions including molecular formulas, number of carbon and hydrogen atoms and number of double bonds were exported to EXCEL. The fragmentation rules and specific peaks were modeled according to the obtained fragmentation patterns. For a series of adducts ($[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M+Ag]^+$ and $[M-H]^-$) specific fragmentations were calculated for each of the lipid classes. The abundances were modeled according to the experimental patterns or values obtained from the literature. Validations were performed using experimental standards, LipidMaps fragmentations and published data from the literature. Around 70,000 MS/MS spectra were the exported into MSP, ASCII and NIST format allowing platform independent use. Using LIB2NIST the library can be used with the NIST MS Search program.

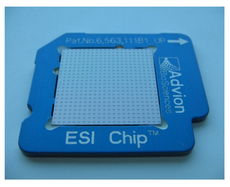
Results

The library can be used in automated or manual library search mode. A typical identification step includes the extraction of the tandem mass spectra, a precursor matching with a specific m/z window and a subsequent MS/MS matching using a dot-product algorithm to create the final hit-score. The resulting hit scores can be used to assign a level of confidence score to each spectrum. The performance of the library was evaluated using external standard compounds and plasma and tissue samples. Using one minute mass spectral infusion experiments with data dependent MS/MS scans around 50 to 80 lipid species can be assigned, with usually ten to twenty compounds with a very high hit sore and thirty to fifty compounds with medium or low score, due to lower signal to noise ratio or interfering or overlapping peaks. Research was supported by R01 ES013932 NIEHS.

Iontrap MS/MS spectra creation

NanoMate nanoESI chip based infusion

Low-resolution LTQ Ion Trap

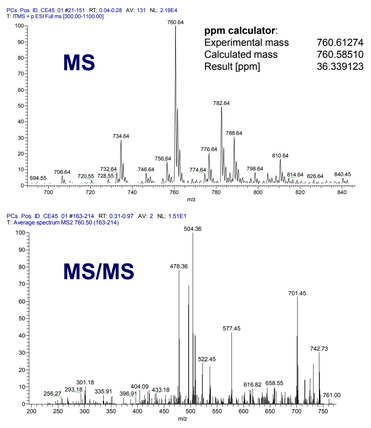


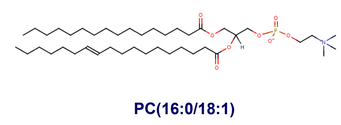
nanoESI chip with 400 nozzles

High-resolution LTQ-FT

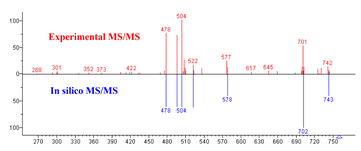
Low-resolution and high-resolution data were obtained from a linear ion trap and a linear iontrap coupled to an Fourier transform (FT-ICR-MS) mass spectrometer. Infusion was performed with a chip-based nanoelectrospray robot (Advion Nanomate). Infusion time per sample with data dependent MS/MS scans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

In-silico tandem mass spectra





PC(16:0/18:1)



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. The double bond position can be determined with MS/MS, but stereochemistry not.

Accurate mass MS and MS/MS

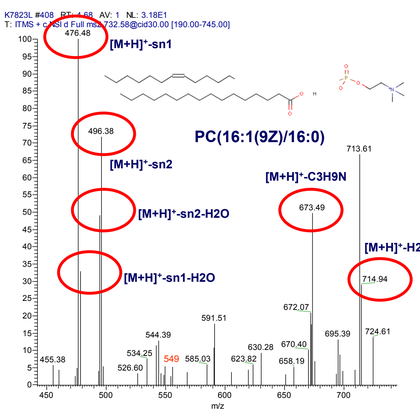
In silico MS/MS

Name: PC 32:1; [M+H]+; GPCho(16:1(9Z)/16:0)
MW: 732.55433
PRECURSORMZ: 732.55433
Num Peaks: 7
714.54377 400 "[M+H]-H2O (-18)"
673.48083 999 "[M+H]-C3H9N (-59)"
549.48829 400 "[M+H]-C5H14NO4P (-183)"
496.34045 600 "[M+H]-sn1"
478.32989 600 "[M+H]-sn1-H2O"
494.32481 600 "[M+H]-sn2"
476.31425 600 "[M+H]-sn2-H2O"

Name: PC 32:1; [M+H]+; GPCho(17:0/15:1(9Z))
MW: 732.55433
PRECURSORMZ: 732.55433
Num Peaks: 7
714.54377 400 "[M+H]-H2O (-18)"
673.48083 999 "[M+H]-C3H9N (-59)"
549.48829 400 "[M+H]-C5H14NO4P (-183)"
480.30917 600 "[M+H]-sn1"
462.29861 600 "[M+H]-sn1-H2O"
510.35609 600 "[M+H]-sn2"
492.34553 600 "[M+H]-sn2-H2O"

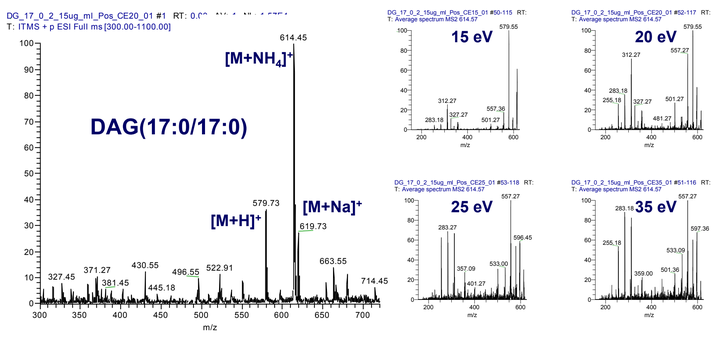
Name: PC 32:1; [M+H]+; GPCho(17:1(9Z)/15:0)
MW: 732.55433
PRECURSORMZ: 732.55433
Num Peaks: 7
714.54377 400 "[M+H]-H2O (-18)"
673.48083 999 "[M+H]-C3H9N (-59)"
549.48829 400 "[M+H]-C5H14NO4P (-183)"
482.32481 600 "[M+H]-sn1"
464.31425 600 "[M+H]-sn1-H2O"
508.34045 600 "[M+H]-sn2"
490.32989 600 "[M+H]-sn2-H2O"

experimental MS/MS



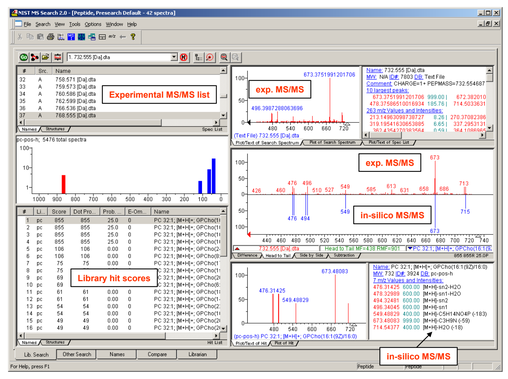
The lipid library contains names with stereoinformation, accurate masses for precursor and tandem spectra as well as information on fragmentations and losses annotated with names and elemental compositions. Conversion in any other library format is possible.

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 such an experiment was performed to obtain optimal fragmentation values.

Manual NIST MS library search



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 can be used for visual inspection of the library hits. 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 unite window) used. After the precursor filter is applied, a dot product hit score is calculated for each spectrum searched.

Lipid species covered in database

Class	Short	Species	Spectra	Positive mode	Negative mode
Glycerophosphocholines	PC	5,476	10,952	M+H; M+Na	low abundant
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	low abundant
Diacylglycerols	DAG	1,764	-	M+NH4; M+H; M+Na	-
Triacylglycerols	TAG	2,640	5,280	M+NH4; M+H; M+Na	-
SUM		**37,075**	**70,622**		

The names of lipid species and their numbers can be found in the table above. Additional adducts including $[M+Li]^+$ or $[M+K]^+$ are sometimes covered in the computational tables, but not exported as library spectra. Each entry needs to be confirmed by experimental or literature evidence. Additional plant lipids such as galactosylglycerides (MGDG, DGDG, SQDG) are not yet covered due to low ionization efficiencies. Additional lipid classes can be easily added via EXCEL template databases.

Example of library search result

Mass [Da]	Class	C:DB	Short name	ESI mode	Full name	Score	Reverse	Probab.
804.500	PC	36:4;	PC 36:4	[M+Na]+	GPCho(20:4(7E,10E,13E,16E)/16:0)	998	998	1.08
874.500	TG	52:3;	TG 52:3	[M+NH4]+	TG(16:0/18:1/18:2)	998	998	99.99
828.500	PC	38:6;	PC 38:6	[M+Na]+	GPCho(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/16:0)	994	994	1.33
758.610	PC	34:2;	PC 34:2	[M+H]+	GPCho(16:0/18:2(2E,4E))	983	983	8.33
808.510	PC	36:2;	PC 36:2	[M+Na]+	GPCho(18:2(9Z,12Z)/18:0)	978	978	0.88
780.517	PC	34:2;	PC 34:2	[M+Na]+	GPCho(18:2(9Z,12Z)/16:0)	973	973	1.69
786.610	PC	36:2;	PC 36:2	[M+H]+	GPCho(18:0/18:2(2E,4E))	972	972	8.33
764.500	PE	38:6;	PE 38:6	[M+H]+	GPEtn(22:5(7Z,10Z,13Z,16Z,19Z)/16:1(9Z))	944	944	1.26
876.500	TG	52:2;	TG 52:2	[M+NH4]+	TG(16:0/18:1/18:1)	940	940	99.99
832.500	PC	38:4;	PC 38:4	[M+Na]+	GPCho(20:4(7E,10E,13E,16E)/18:0)	918	918	1.21
846.500	TG	50:3;	TG 50:3	[M+NH4]+	TG(16:0/16:1/18:2)	878	878	99.99
872.500	TG	52:4;	TG 52:4	[M+NH4]+	TG(16:0/18:2/18:2)	842	842	73.74
782.550	PC	34:1;	PC 34:1	[M+Na]+	GPCho(18:1(9Z)/16:0)	800	800	3.13
848.500	TG	50:2;	TG 50:2	[M+NH4]+	TG(16:0/16:1/18:1)	775	775	97.2
716.500	PE	34:2;	PE 34:2	[M+H]+	GPEtn(20:1(13Z)/14:1(9Z))	700	700	1.65
856.500	PC	40:6;	PC 40:6	[M+Na]+	GPCho(20:5(5Z,8Z,11Z,14Z,17Z)/20:1(13Z))	614	614	2.13
898.500	TG	54:5;	TG 54:5	[M+NH4]+	TG(18:1/18:2/18:2)	572	572	92.16
760.620	PC	34:1;	PC 34:1	[M+H]+	GPCho(16:0/18:1(11E))	553	553	5.55
870.500	TG	52:5;	TG 52:5	[M+NH4]+	TG(16:0/18:2/18:3)	511	511	96.61
792.500	PE	40:6;	PE 40:6	[M+H]+	GPEtn(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/18:0)	468	468	1.47
518.500	PC	15:0;	PC 15:0	[M+Na]+	GPCho(12:0/3:0)	435	435	21.42
810.520	PC	36:1;	PC 36:1	[M+Na]+	GPCho(18:1(9Z)/18:0)	413	413	3.64
896.500	TG	54:6;	TG 54:6	[M+NH4]+	TG(16:0/18:2/20:4)	410	410	98.02
788.500	PC	36:1;	PC 36:1	[M+H]+	GPCho(18:0/18:1(11E))	397	397	5.52
809.600	SM	40:1;	SM 40:1	[M+Na]+	SM(d14:0/26:1(17Z))	390	390	19.99
756.500	PC	32:0;	PC 32:0	[M+Na]+	GPCho(16:0/16:0)	387	387	21.62
756.500	PC	32:0;	PC 32:0	[M+Na]+	GPCho(19:0/13:0)	329	329	7.88
836.500	PC	38:2;	PC 38:2	[M+Na]+	GPCho(24:1(15Z)/14:1(9Z))	84	84	1.1
836.500	PC	38:2;	PC 38:2	[M+Na]+	GPCho(24:1(15Z)/14:1(9Z))	84	84	1.01
756.500	PC	32:0;	PC 32:0	[M+Na]+	GPCho(19:0/13:0)	82	82	2.36

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 below 100 contains possibly wrong identifications.

Conclusions

- The MS/MS library contains 70,000 mass spectra of 37,000 lipids in negative and positive mode and covers multiple adducts
- Library is freely available in open ASCII, MSP and NIST format
- Phospholipids and triacylglycerols can be identified from LC-MS/MS or infusion MS/MS using traditional library search with precursor filtering and NIST dot product MS/MS matching
- Low resolution iontrap technology together with MS/MS scans used for phospholipid identification via head group and neutral loss analysis of fatty acids
- Tetrahedral (R/S) and double bond (E/Z) stereochemistry can not be determined with MS/MS only. The correct determination of the double bond position is possible.
- Chip-based nanoESI infusion allows fast acquisition (one minute per sample) in high/low resolution mode