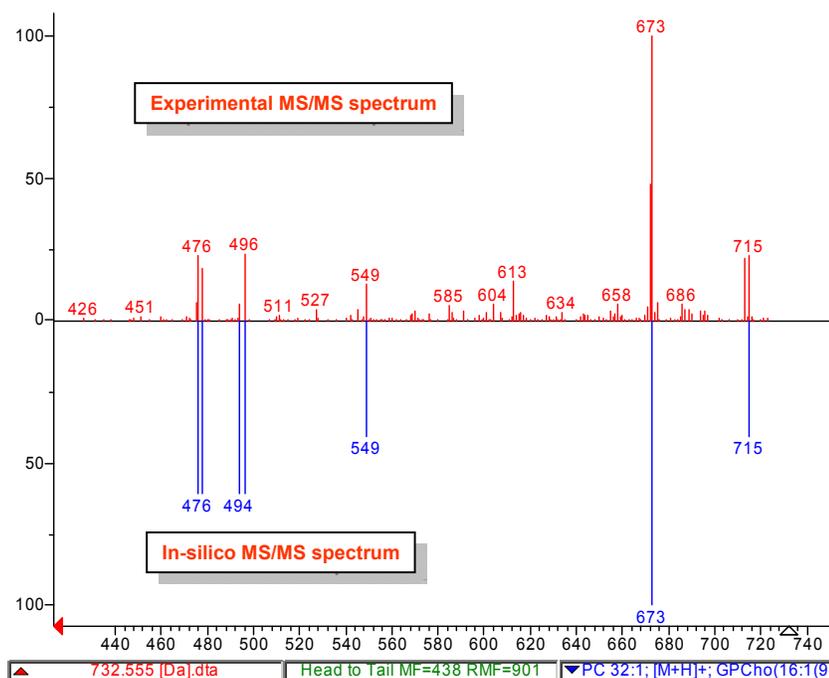


User Manual



Lipid identification using the LipidBlast libraries



Tobias Kind and Oliver Fiehn

Fiehnlab November 2012 v46

Contact: Dr. Tobias Kind - <http://fiehnlab.ucdavis.edu/staff/kind/>

General Disclaimer: The following information is provided for research purposes only and may contain errors.

LipidBlast Introduction

What is LipidBlast?

LipidBlast is a collection of in-silico electrospray tandem mass spectral libraries for the identification of neutral and polar lipid species. In the current development stage around 120,000 MS/MS spectra, covering multiple glycerophospholipid classes and possible ESI adducts are included. For more resources regarding lipids and analytical techniques please visit:

- [LipidMaps](#)
- [Lipid Analysis with GC-MS, LC-MS, FT-MS - Metabolomics Fiehn Lab](#)

The LipidBlast libraries were developed from known theoretical fragmentations and experimental fragmentation and abundance information from tandem mass spectra (MS/MS). The experimental basis was a Thermo Finnigan LTQ linear iontrap instrument equipped with an Advion Nanomate nanoESI robotic infusion system. However the libraries are platform independent and can be used for all MS/MS or MS^E platforms from all mass spectrometry vendors including ABI Sciex, Agilent, Bruker, Shimadzu, Thermo, Waters and others.

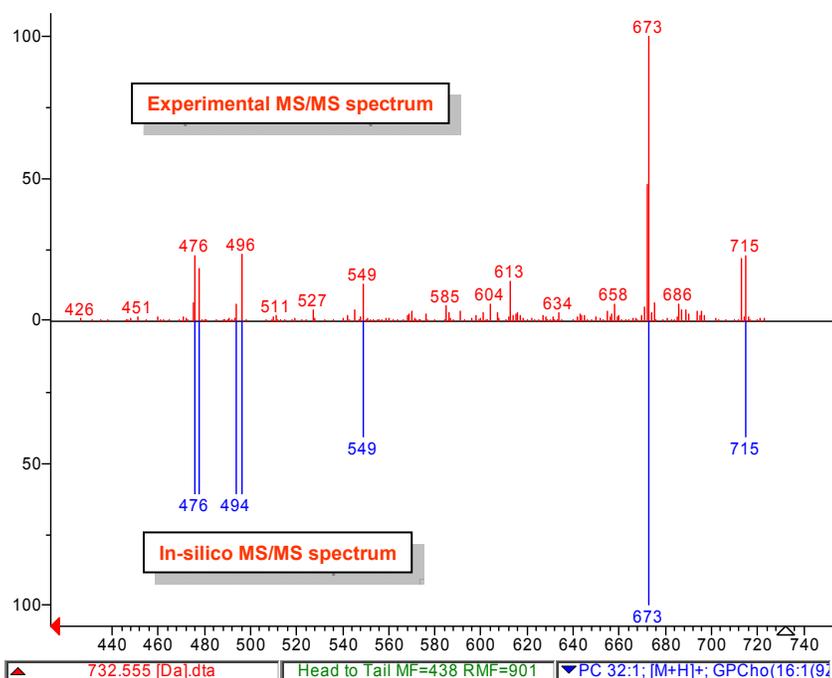


Figure: LipidBlast libraries showing an experimental MS/MS spectrum and a matching computer generated in-silico mass spectrum with similar fragment peaks and abundances.

How can the LipidBlast libraries be used?

The LipidBlast can be used for **identification of polar lipids** using high resolution and low resolution tandem mass spectrometer data. In any case the precursor mass is needed from the MS¹ spectrum and the data-dependent MS² spectra must be extracted from the experimental runs. Therefore the LipidBlast libraries can be used for identification of lipids from:

- A) **Infusion mass spectrometric** experiments with data dependent MS/MS scans
- B) **LC-MS experiments** with data dependent MS/MS scans
- C) **LC-MS experiments** with TOF MS^E experiments

The LipidBlast libraries are traditional mass spectral libraries. One difference to other in-silico libraries is that accurate mass information is included for all precursor and product ions, as well as fragment names and modeled abundances. The following mass spectral search modes can be applied:

- A) **Traditional mass spectral library search**
- B) **MS/MS database search**

The LipidBlast libraries can be used with traditional library search procedures such as PBM search or dot product (NIST MS) search. Better performance and better hit rates and higher identification rates are obtained if the accurate precursor mass is included (for pre-filtering possible lipid species) and a subsequent MS/MS peak matching is performed. The NIST MS Search algorithm is capable of finding a single mass spectrum in a library of one million spectra in only milliseconds.

How can I extract MS/MS spectra?

If your platform software has MS/MS search capabilities, just import the provided ASCII MSP spectra into your vendor software. If your platform has no MS/MS search capability (which includes precursor search and product ion search step, not to be confused with normal mass spectral library search) you can extract MS/MS spectra, export them as MFG files (MASCOT format) and import them directly into NIST MS Search or NIST MSPepSearch. These approaches are used in the proteomics community since many years, so talking to a fellow proteomics researcher will solve that problem.

- A) For extraction of MGF directly from raw files, download ProteoWizard
<http://proteowizard.sourceforge.net/>
- B) Other software can be found at the Sashimi tools
[\(http://sourceforge.net/projects/sashimi/files/](http://sourceforge.net/projects/sashimi/files/)
- C) Trans Proteomic Pipeline software page
<http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML>).

How can I perform quantification of lipid species?

The LipidBlast library is a mass spectral library for identification of lipids. In order to quantify lipids other commonly available vendor software tools must be used. This includes the use of internal or external standards (reference compounds), the use of proper calibration and peak integration methods and programs for alignment of LC-MS or direct infusion MS data. For a collection of programs see:

<http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/LipidAnalysis/>

For alignment programs see peak alignment of LC, GC, MS, NMR data

http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak_Alignment/

Or see our poster using the Genedata Refiner MS program: An automated workflow for rapid alignment and identification of lipid biomarkers obtained from chip-based direct infusion nanoelectrospray tandem mass spectrometry

<http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/LipidAnalysis/lipids-biomarker-workflow.pdf>

Which lipid classes are covered in the LipidBlast libraries?

Currently the glycerophospholipids are available in NIST MS MSP format. Other libraries (monoacylglycerols, diacylglycerols and triacylglycerols) and galactolipids (MGDG, DGDG, SQDG) are also available. Additional libraries can be developed on request or by using the provided data templates. The creation of in-silico libraries depends on the correct molecular species enumeration and existence of molecular structures and secondly on a correct nomenclature to help during the discovery process. Secondly experimental spectra from reference standards must be obtained for all possible classes all possible side chain modifications. That includes symmetric sn1 and sn2 side chains, different degrees of unsaturation and different head groups.

Table: LipidBlast lipids and current adduct libraries including lyso-forms for some species

Class	Short	Positive mode	Negative mode
Glycerophosphocholines	PC	M+H; M+Na	low abundant
Glycerophosphoethanolamines	PE	M+H; M+Na	M-H
Glycerophosphoserines	PS	M+H	M-H
Glycerophosphoglycerols	PG	low abundant	M-H
Glycerophosphoinositols	PI	low abundant	M-H
Glycerophosphates	PA	M+NH ₄ ; M+Na; M+2Na	M-H
Sphingomyelins	SM	M+H; M+Na	low abundant
Diacylglycerols	DAG	M+NH ₄ ; M+H; M+Na	-
Triacylglycerols	TAG	M+NH ₄ ; M+H; M+Na	-
Monogalactosyldiacylglycerols	MGDG	M+NH ₄ ; M+Na	M-H
Digalactosyldiacylglycerols	DGDG	M+NH ₄ ; M+Na	M-H
Sulfoquinovosyldiacylglycerols	SQDG		M-H
Gangliosides	[glycan]-Cer		M-H
Ceramide-1-phosphates	CerP	M+H	M-H
Sulfatides	ST		M-H
Gangliosides	[glycan]-Cer		M-H
Diacylated phosphatidylinositol monomannoside	Ac2PIM1		M-H
Diacylated phosphatidylinositol dimannoside	Ac2PIM2		M-H
Triacylated phosphatidylinositol dimannoside	Ac3PIM2		M-H
Tetraacylated phosphatidylinositol dimannoside	Ac4PIM2		M-H
Sulfatides	ST		M-H

Some of the adducts may not occur in the literature, because ions may be suppressed in mixtures, but ions were observed from authentic standards. Also common adducts such as [M+Li]⁺ may not

be included because experimental spectra were not acquired and in-silico spectra were not modeled yet.

How large is each single adduct library? The size depends on the lipid species but is usually around 5,500 in-silico MS/MS spectra. In case of sphingomyelins the library size is smaller. So at the current development stage around 212,516 MS/MS spectra from 119,200 structures are included. In case of TAGs the library size was reduced to the 2640 most common compounds (TG(12:0/12:0/16:0) to TG(22:6/22:6/22:6)) due to the large number of combinations.

NIST MS Search program for easy identification

The NIST MS Search GUI is a freely available program developed by the Mass Spectral Data group at NIST. The procedure of lipid identification, can be easily performed by loading the result *.mgf file into the NIST MS GUI search. By clicking the green GO button or each single experimental MS/MS spectrum a library search is performed. Hit scores are provided in the left result window. See <http://chemdata.nist.gov/> and the NIST help file (Ver20Man.pdf) for further information.

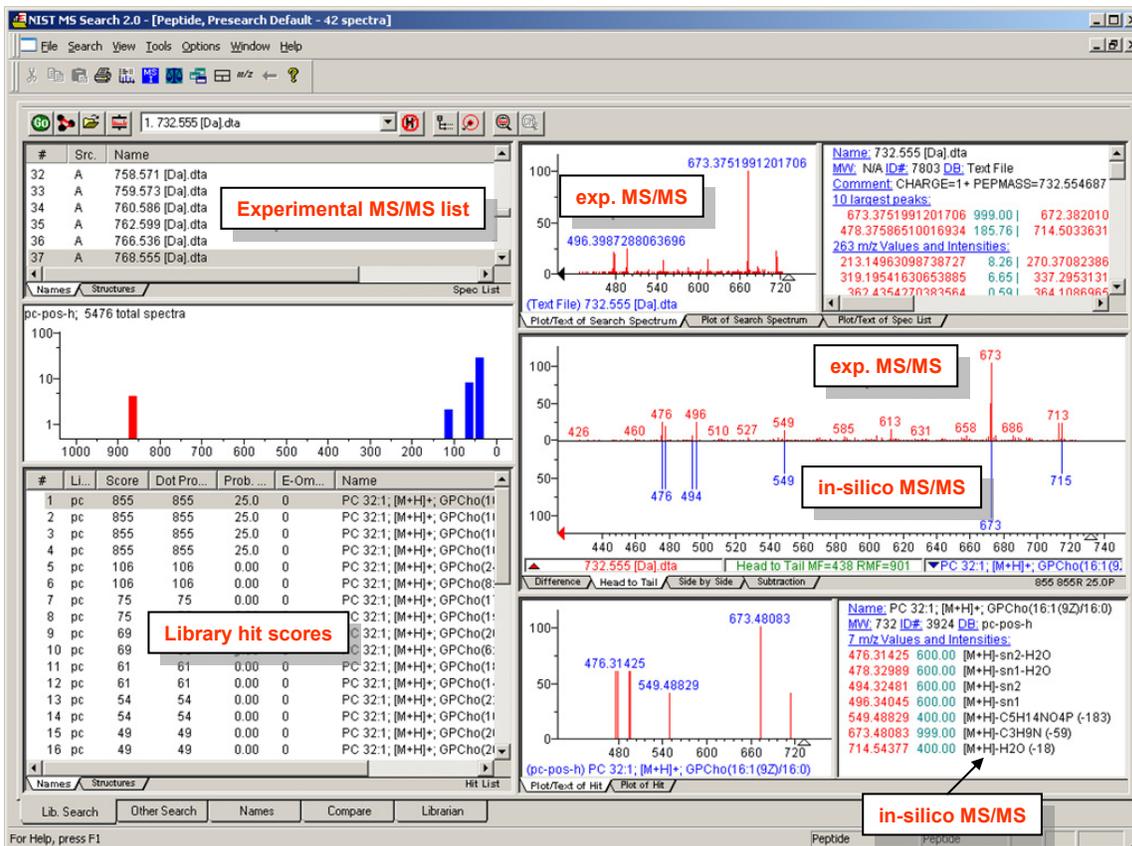
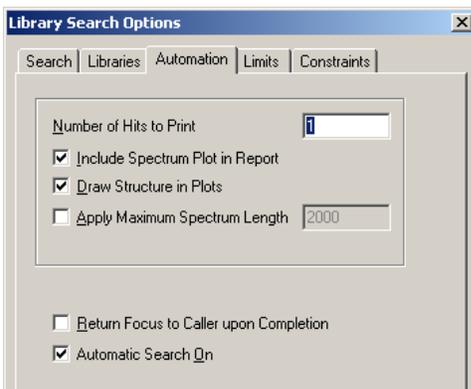


Figure: NIST MS search GUI with the LipidBlast libraries loaded



Automation of library search is obtained by starting nistms.exe with the /par=4 option from the DOS command prompt which writes all results into NISTLOG.txt in the same directory. The option nistms.exe /par=8 stops the automated result logging. The automation options are set in the library options dialog. The number of hits is the most important selection.

Library search options and m/z tolerances

The NIST MS Search option MS/MS search has to be selected as seen in the picture. An MS/MS search precursor mass tolerance of 0.001 Da or in case of accurate mass data and 0.4 m/z units for unit mass spectrometers should be selected. The product ion mass can be less stringent, due to hydrogen shifts, 0.4 to 0.8 Da is acceptable.

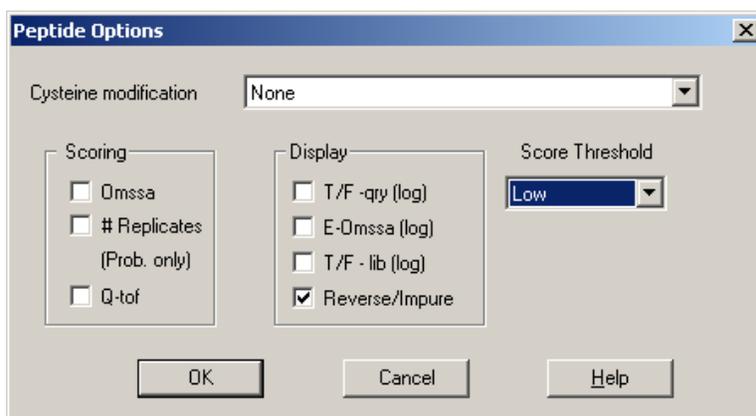
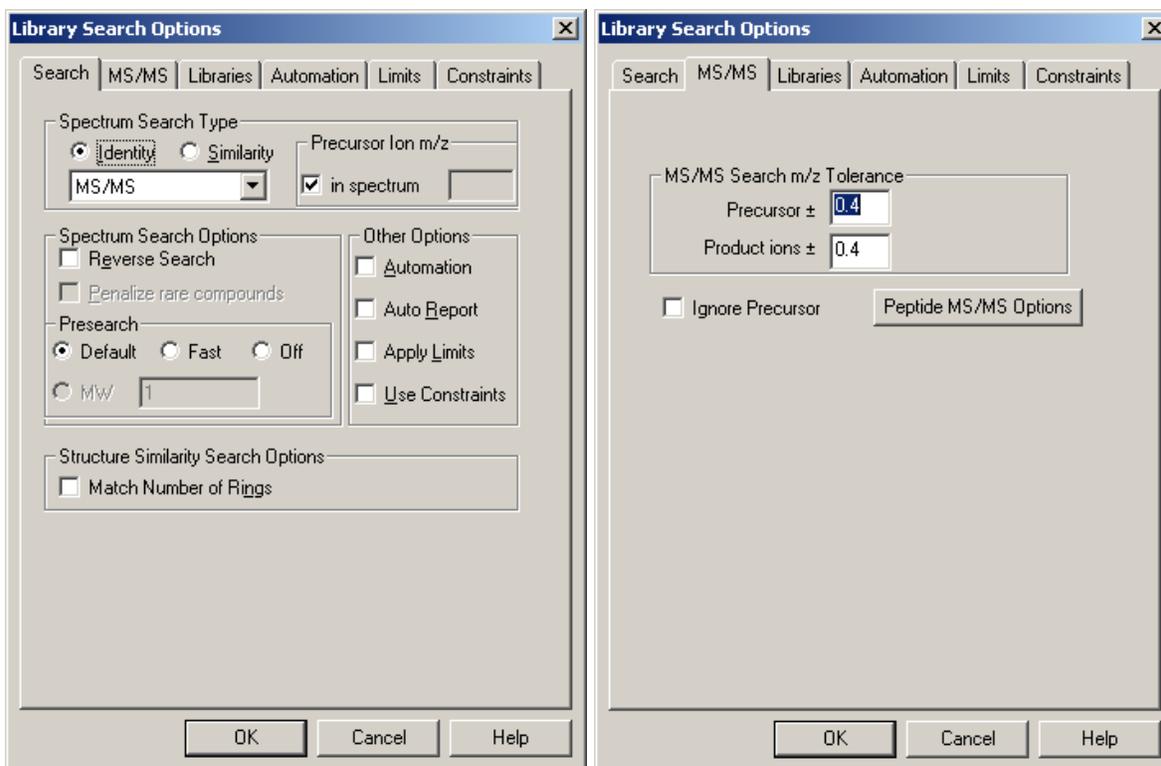
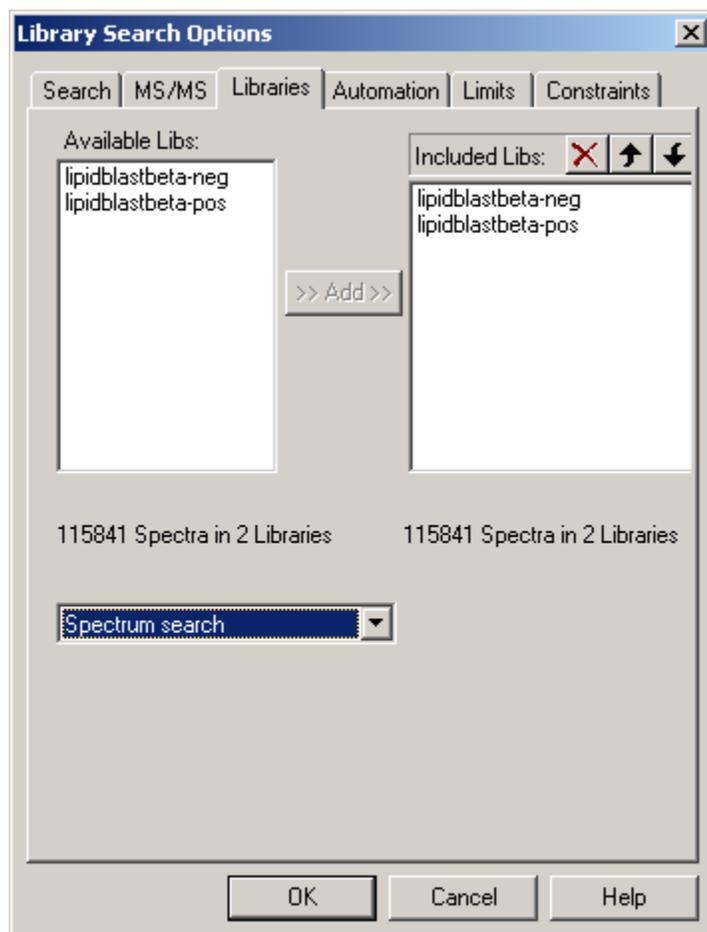


Figure: MS/MS search and appropriate m/z windows must be selected. MS/MS options are selected using reverse search and low score thresholds, all other options are off.

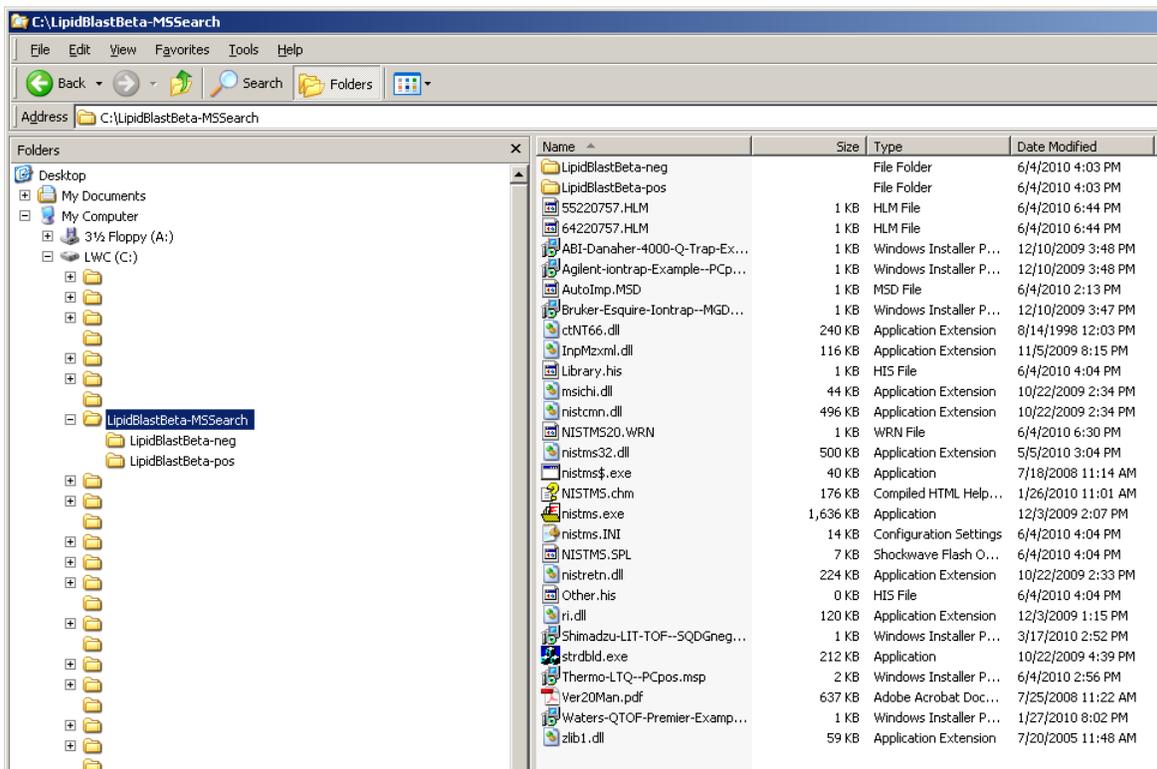
Library selection and install options

The libraries must be activated and visible in the right window. If no library is selected make sure to edit the install directory (NISTMS.ini) or simply delete the NISTMS.ini file and restart nistms.exe. The program will create a new INI file and will ask for the current library directory.



Installation NIST MS Search + LipidBlast procedure step-by-step

The install assumes a fully patched Windows OS (Vista, XP, Server). Extract the directories from the zip file and place it under C:\ for instant use, or install into any subdirectory of your choice (requires selection of library path). The folder should be C:\LipidBlastBeta-MSSearch\ Start nistms.exe to work with the program.



(1) The NIST MS Search GUI

The NIST MS Search GUI is a freely available program developed by the Mass Spectral Data group at NIST. The current distribution of the ZIP file contains a working version of the NIST MS Search GUI program. See <http://chemdata.nist.gov/> and the NIST help file (Ver20Man.pdf) for further information.

(2) The LipidBlast libraries

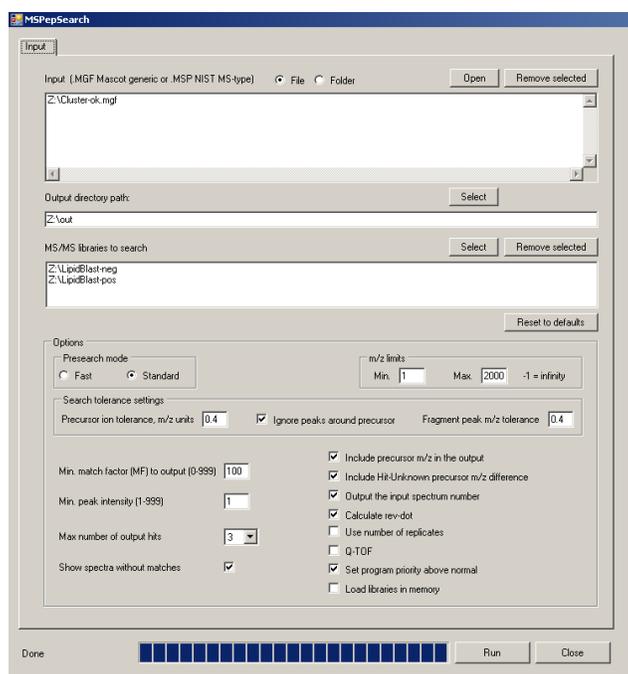
The LipidBlast libraries are developed by Tobias Kind, Kwang-Hyeon Liu, Do Yup Lee and Oliver Fiehn; Fiehnlab June 2010 v31 <http://fiehnlab.ucdavis.edu/staff/kind/>

(1) and (2) are independent developments and ***not*** associated with each other.

Using LipidBlast in high-throughput mode

LipidBlast can be used to annotate lipids from large MS/MS files with the NIST MSPepSearch. The search speed is several hundred spectra per second and thousands of spectra can be annotated at-once. (1) Import the MGF file on the top window (2) Select an output drive (4) Select the LipidBlast directories from the prior GUI installation such as C:\LipidBlastBeta-MSSearch\

Download here: http://peptide.nist.gov/software/ms_pep_search_gui/MSPepSearch.html



Define precursor and product ion accuracy, select number of output hits, do ***not*** load into memory (current bug that prevents libraries searched) and press run. An *.tsv file with library search scores will be created which can be imported into EXCEL or OpenOffice Calc.

2	3	4	Num	Unknown	Precursor m/z	Rank	Library	Id	Mass	Delta(m/z)	Lib	Precursor	Score	Dot	Product	Prob(%)	Rev-Dot	Peptide
71	38	734.566	[Da]	734.566	3	LipidBlast-pos	43806	734	-0.1	734.47	274	342	1.6	953	PC 31:4,	[M+Na] ⁺ ,	GPCho(8:0/22:4(7Z,10Z,13Z,16Z))	
72	39	734.623	[Da]	734.623	1	LipidBlast-pos	43807	734	-0.15	734.47	331	468	10.1	949	PC 31:4,	[M+Na] ⁺ ,	GPCho(11:0/20:4(6E,8E,11E,14E))	
73	39	734.623	[Da]	734.623	2	LipidBlast-pos	43808	734	-0.15	734.47	331	468	10.1	949	PC 31:4,	[M+Na] ⁺ ,	GPCho(11:0/20:4(6Z,8Z,11Z,14Z))	
74	39	734.623	[Da]	734.623	3	LipidBlast-pos	43809	734	-0.15	734.47	331	468	10.1	949	PC 31:4,	[M+Na] ⁺ ,	GPCho(11:0/20:4(7E,10E,13E,16E))	
75	40	736.500	[Da]	736.5	1	LipidBlast-pos	38614	736	0	736.49	160	187	50	653	PC 33:6,	[M+H] ⁺ ,	GPCho(11:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	
76	40	736.500	[Da]	736.5	2	LipidBlast-pos	38615	736	0	736.49	160	187	50	653	PC 33:6,	[M+H] ⁺ ,	GPCho(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/11:0)	
77	41	736.500	[Da]	736.5														
78	42	742.500	[Da]	742.5	1	LipidBlast-pos	43904	742	0	742.54	345	508	6.2	948	PC 31:0,	[M+Na] ⁺ ,	GPCho(6:0/25:0)	
79	42	742.500	[Da]	742.5	2	LipidBlast-pos	43923	742	0	742.54	345	508	6.2	948	PC 31:0,	[M+Na] ⁺ ,	GPCho(25:0/6:0)	
80	42	742.500	[Da]	742.5	3	LipidBlast-pos	43903	742	0	742.54	334	496	4.2	947	PC 31:0,	[M+Na] ⁺ ,	GPCho(5:0/26:0)	
81	43	746.500	[Da]	746.5	1	LipidBlast-pos	43929	746	0	746.47	289	404	11	931	PC 32:5,	[M+Na] ⁺ ,	GPCho(12:0/20:5(6Z,8Z,11Z,14Z,17Z))	
82	43	746.500	[Da]	746.5	2	LipidBlast-pos	43934	746	0	746.47	289	404	11	931	PC 32:5,	[M+Na] ⁺ ,	GPCho(20:5(6Z,8Z,11Z,14Z,17Z)/12:0)	
83	43	746.500	[Da]	746.5	3	LipidBlast-pos	43927	746	0	746.47	279	392	7.8	919	PC 32:5,	[M+Na] ⁺ ,	GPCho(10:0/22:5(4Z,7Z,10Z,13Z,16Z))	
84	44	746.500	[Da]	746.5	1	LipidBlast-pos	38734	746	0.1	746.57	117	156	5.7	672	PC 33:1,	[M+H] ⁺ ,	GPCho(16:0/17:1(8Z))	
85	44	746.500	[Da]	746.5	2	LipidBlast-pos	38739	746	0.1	746.57	117	156	5.7	672	PC 33:1,	[M+H] ⁺ ,	GPCho(17:1(8Z)/16:0)	
86	44	746.500	[Da]	746.5	3	LipidBlast-pos	38724	746	0.1	746.57	113	151	4.8	670	PC 33:1,	[M+H] ⁺ ,	GPCho(15:0/18:1(11E))	
87	45	748.500	[Da]	748.5	1	LipidBlast-pos	43937	748	0	748.49	271	343	6.6	930	PC 32:4,	[M+Na] ⁺ ,	GPCho(8:0/24:4(6Z,8Z,11Z,14Z))	
88	45	748.500	[Da]	748.5	2	LipidBlast-pos	43954	748	0	748.49	271	343	6.6	930	PC 32:4,	[M+Na] ⁺ ,	GPCho(24:4(6Z,8Z,11Z,14Z)/8:0)	
89	45	748.500	[Da]	748.5	3	LipidBlast-pos	43944	748	0	748.49	270	342	6.4	917	PC 32:4,	[M+Na] ⁺ ,	GPCho(14:1(8Z)/18:3(6Z,9Z,12Z))	
90	46	750.500	[Da]	750.5														
91	47	754.500	[Da]	754.5	1	LipidBlast-pos	44046	754	0	754.54	423	625	3.7	974	PC 32:1,	[M+Na] ⁺ ,	GPCho(16:0/16:1(7Z))	
92	47	754.500	[Da]	754.5	2	LipidBlast-pos	44047	754	0	754.54	423	625	3.7	974	PC 32:1,	[M+Na] ⁺ ,	GPCho(16:0/16:1(8Z))	
93	47	754.500	[Da]	754.5	3	LipidBlast-pos	44048	754	0	754.54	423	625	3.7	974	PC 32:1,	[M+Na] ⁺ ,	GPCho(16:1(7Z)/16:0)	
94	48	756.500	[Da]	756.5	1	LipidBlast-pos	44079	756	0	756.55	400	640	21.4	987	PC 32:0,	[M+Na] ⁺ ,	GPCho(16:0/16:0)	
95	48	756.500	[Da]	756.5	2	LipidBlast-pos	44077	756	0	756.55	370	611	6	961	PC 32:0,	[M+Na] ⁺ ,	GPCho(14:0/18:0)	
96	48	756.500	[Da]	756.5	3	LipidBlast-pos	44081	756	0	756.55	370	611	6	961	PC 32:0,	[M+Na] ⁺ ,	GPCho(18:0/14:0)	
97	49	758.500	[Da]	758.5	1	LipidBlast-pos	44090	758	0	758.47	322	479	23.8	943	PC 33:6,	[M+Na] ⁺ ,	GPCho(11:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	

Detailed technical information

Installation of the LipidBlast libraries using a Virtual Machine

Technical Recommendations

The LipidBlast libraries require the installation of multiple software programs. In order to comply with industry computer safety standards and for testing purposes it is recommended to install all programs into a Virtual Machine. Some of the programs are not compatible with older Microsoft .NET (dot net) technology, so all required updates must be installed.

Visit [Virtual machine - Wikipedia, the free encyclopedia](#) to learn more about Virtual Machines. For ease of use we use the freely Microsoft Virtual PC and install a licensed (purchase required) operating system or test license from Microsoft for Windows XP, Windows Vista or Windows Server. Universities can obtain a research license for VMWare Workstation or a free Sun Virtual Box license.

Free Sun Virtual BOX:

www.virtualbox.org

Free Microsoft Virtual PC 2007:

[Microsoft Virtual PC 2007](#)

OS Downloads:

[Microsoft Windows Server 2003 R2 Enterprise Edition VHD](#)

[Windows Server 2008 Evaluation Virtual Hard Drive Images](#)

ZIP files:

[7Zip program for UNZIP operations of packed files](#) (zip/gz)

Follow the recommendations and install the Microsoft Virtual PC 2007 and install any licensed Windows XP, Windows Vista or Windows Server operating system into the virtual machine. Each new virtual environment requires a single OS license (unless used under WIN7). Read more in the following publication:

Software platform virtualization in chemistry research and university teaching

Tobias Kind , Tim Leamy , Julie A Leary and Oliver Fiehn

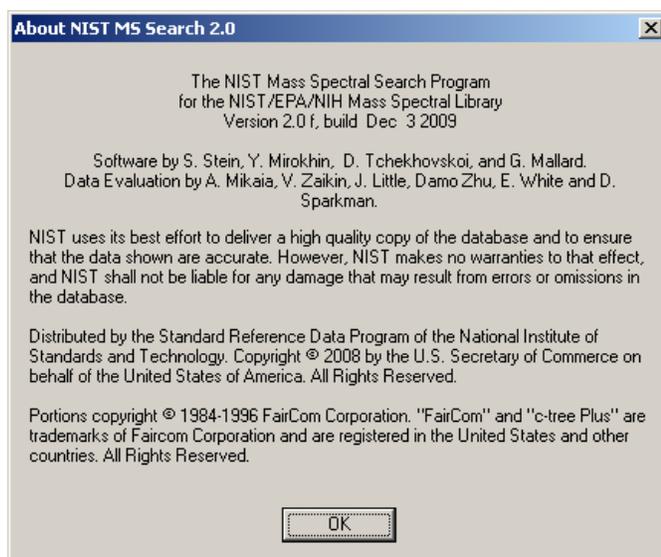
Journal of Cheminformatics 2009, 1:18doi:10.1186/1758-2946-1-18

<http://www.jcheminf.com/content/1/1/18>

Useful software

The LipidBlast libraries can be utilized with multiple software packages and workflows. We suggest using the freely available and latest version of the NIST MS Search GUI (compiled in May 2010). For technical questions please refer to the PDF Handbook (Ver20Man.pdf) - NIST Mass Spectral Search Program (Version 2.0f) provided by the NIST Mass Spectrometry Data Center

NIST MS Search GUI at <http://chemdata.nist.gov/>



NIST MSPepSearch for high-throughput annotation

- Software for batch interpretation of tandem mass spectra. Especially useful for thousands of tandem mass spectra, evaluations can be around 1500 spectra / second
http://peptide.nist.gov/software/ms_pep_search_gui/MSPepSearch.html

Sashimi file conversion utilities

- Tools for conversion of MS/MS spectra to mzXML and subsequent *.dta and *.mgf conversion can be found at <http://proteowizard.sourceforge.net/> and <http://sashimi.sourceforge.net/>

MS-Clustering

- The clustering tool is required to improve signal to noise ratio from infusion mass spectra. Similar MS/MS spectra are clustered according to selected precursor accuracy and a single consensus spectrum is created. Mscluster.exe accepts several formats, among them dta, mgf, mzXML. For LC-MS data a retention time tolerance match may be required or true mass spectral deconvolution, because peak data at different retention times should **not** be clustered. MSCluster can be downloaded from <http://peptide.ucsd.edu/> or <http://proteomics.ucsd.edu/Software/MSClustering.html>

Thermo extract_msn.exe for dta file extraction

- This tool is part of Thermo Xcalibur 2.0 and used for extraction of *.dta files from data dependent scans and for the creation of *.mgf files (combined *.dta files). A license must be obtained from Thermo Fisher Inc.

DeconMSn - optional tool for dta and mgf extraction

- DeconMSn is a tool for dta and mgf file extraction from PNNL. It requires the installation of the Thermo XCalibur software and multiple Microsoft dot NET upgrades. DeconMSn can be downloaded at <http://omics.pnl.gov/>

DTASuperCharge - optional tool for dta and mgf extraction

- DTASuperCharge is a GUI driven tool for MGF file creation and can be obtained from the following server <http://msquant.sourceforge.net/>

mgfprecursor.exe – tool for precursor ion annotation

- After the clustering algorithm spectra contain only cluster IDs. For the NIST MS Peptide Search it is more convenient to have attached m/z precursor values. That also allows better evaluation during automation issues. The tool requires the MS Virtual C++ V2008 redistributable package installed as well as all Microsoft dot NET updates. The tool is provided as source in C for Visual C++ and also as executable.

create-clustererd-mgf.bat – automatic processing of Thermo *.raw files

- This Windows batch files automates the whole process of a) creating dta files b) clustering similar spectra c) rewriting the clustered m/z precursor ions e) creating the mgf file which can be imported into NIST MS Peptide search. This batch file uses **DeconMSn.exe** as MGF file creator

create-cluster-mfg.bat – automatic processing of Thermo *.raw files

- This Windows batch files automates the whole process of a) creating dta files b) clustering similar spectra c) rewriting the clustered m/z precursor ions e) creating the mgf file which can be imported into NIST MS Peptide search. This batch file uses **extract_msn.exe** as DTA file creator and MGF files must be created independently or by MSCluster.

Visual C++ 2008 Redistributable Package (x86)

- Required for C++ applications can be downloaded from [Microsoft](#) vcredist_x86.exe

Nomenclature changes for glycerophospholipids

Please be aware that the nomenclature changed so now the short term abbreviations are used in all LipidBlast libraries with the long version in the explanation. For current updates see online.

LipidMaps Updates:

http://www.lipidmaps.org/data/classification/classification_updates.html

Table: Changes in nomenclature

Class	Previous	New
Glycerophosphocholines	GPCho	PC
Glycerophosphoethanolamines	GPEtn	PE
Glycerophosphoserines	GPSer	PS
Glycerophosphoglycerols	GPGro	PG
Glycerophosphoglycerophosphates	GPGroP	PGP
Glycerophosphoinositols	GPIns	PI
Glycerophosphoinositol monophosphates	GPInsP	PIP
Glycerophosphoinositol bis-phosphates	GPInsP2	PIP2
Glycerophosphoinositol tris-phosphates	GPInsP3	PIP3
Glycerophosphates	GPA	PA
Glyceropyrophosphates	GPP	PPA
Glycerophosphoglycerophosphoglycerols	CL	CL
CDP-glycerols	GCDP	CDP-DG
Glycosylglycerophospholipids	[glycan]-GP	[glycan]-GP
Glycerophosphoinositolglycans	[glycan]-GPIns	[glycan]-PI
Glycerophosphonocholines	GPnCho	PnC
Glycerophosphonoethanolamines	GPnEtn	PnE

How was the LipidBlast library created?

The fragmentations were obtained from the literature and LipidMaps data and validated using external reference compounds standards. Those reference compounds were investigated using different ionization voltages and using multiple solvents.

Short Nam	Neutral				POS			
	SNI acyl loss	SNI-H2O	SN2 acyl loss	SN2-H2O	[M+H] ⁺ Pre	fragment -H2O	fragment -C3H9N	fragment -C5H14ND
0 PC 36.4	495.3326	477.3221	543.3326	525.3221	782.5700	764.5594	723.4965	599.5039
0 PC 36.4	467.3013	449.2908	571.3639	553.3533	782.5700	764.5594	723.4965	599.5039
0 PC 36.4	439.2701	421.2595	599.3952	581.3846	782.5700	764.5594	723.4965	599.5039
2 PC 36.3	575.3952	557.3847	465.2857	447.2752	784.5856	766.5751	725.5121	601.5196
3 PC 36.3	545.3483	527.3377	495.3327	477.3221	784.5856	766.5751	725.5121	601.5196
3 PC 36.3	545.3483	527.3377	495.3327	477.3221	784.5856	766.5751	725.5121	601.5196
2 PC 36.3	547.3639	529.3534	493.3170	475.3065	784.5856	766.5751	725.5121	601.5196
2 PC 36.3	547.3639	529.3534	493.3170	475.3065	784.5856	766.5751	725.5121	601.5196
2 PC 36.3	547.3639	529.3534	493.3170	475.3065	784.5856	766.5751	725.5121	601.5196
2 PC 36.3	547.3639	529.3534	493.3170	475.3065	784.5856	766.5751	725.5121	601.5196
3 PC 36.3	517.3170	499.3065	523.3639	505.3534	784.5856	766.5751	725.5121	601.5196

Figure: The LipidBlast libraries were developed from EXCEL templates containing all highly probable fragmentations.

In principle every possible electrospray (ESI) or atmospheric pressure chemical ionization (APCI) adduct can be modeled. Currently $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M-H]^-$ for both positive and negative ESI are covered. Special iontrap fragmentations not mentioned in the literature before are also covered. Intensities can be modeled according to iontrap or QTOF data and different CID ionization potentials. The current libraries are developed using 25V CID as reference for peak abundance modeling.

```
Sub exportMSMS2mepPC()
'Program for converting MS/MS fragmentation data into MSP files
'for NIST Peptide search
'written by Tobias Kind (2009) fiehnlab.ucdavis.edu
'License (CC-by) Creative commons by attribution
'commercial and non-commercial use allowed

Dim PrintString As String
Dim extMSPName As String
Dim MSPNameString As String

'Start with GlobalMassMin itself (CLASSChange ALL)
For i = 1 To 5476 Step 1
    '-----
    'Name: KKEEELVALK/3
    PrintString = "Name: " & Range("X" & 6 + i) & "; " & "[M+H]+" & "; " & F
    Print #FileNumWrite, PrintString

    'MW: 1317 -
    PrintString = "MW: " & RoundFixed(Range("AC" & 6 + i), RoundingAccuracy,
    Print #FileNumWrite, PrintString

    'PRECURSORMZ:
    PrintString = "PRECURSORMZ: " & RoundFixed(Range("AC" & 6 + i), Rounding
    Print #FileNumWrite, PrintString

    'Comments: put all in for copy/paste
    PrintString = "Comments: " & Range("X" & 6 + i) & "; " & "[M+H]+" & "; F
    Print #FileNumWrite, PrintString

```

Figure: Around 2000 lines of EXCEL VBA code were used for MSP export.

Automated generation of clustered MGF files for Thermo XCalibur

The install assumes a clean totally patched 32-bit Windows OS (Vista, XP, Server). The current procedure describes the extraction of MS/MS data as dta files. Please refer to the Internets for more information about DTA, MGF, PKL files. Those formats are very common in the proteomics community and there exist many web resources. The clustered MGF files are important because they will be imported into the NIST MS Search program. MGF files usually contain multiple MS/MS scans. The current procedure describes the installation for Thermo Finnigan XCalibur *.RAW format. For other vendors please refer to web or vendor information.

- 1) Install XCalibur 2.0 with all service packs.
- 2) Create a directory C:\MSMS\
Create a directory C:\MSMS\work\
Create a directory C:\MSMS\MSCluster
Copy create-cluster-mfg.bat into the work directory
Copy mgfprecursor.exe into the work directory; The files UnifiedFile.dll; Fregistry.dll and Fileio.dll are part of XCalibur and are needed in the directory unless no PATH is set
Copy extract_msn.exe from the XCalibur install into the C:\MSMS\work\ directory
Copy PCs_Pos_ID_CE45_01.raw as example file into the C:\MSMS\work\ directory

When using DeconMSn.exe -XMGF the MGF file is directly created by DeconMSn. In this case please use create-clustererd-mgf.bat. Otherwise the DTA file have to be bundled into an DTA file independently. In case of problems open the BATCH file (Copy create-cluster-mfg.bat) which contains a documented structure of all calls which will be executed.

- 3) Download any Windows compatible Perl distribution (AtiveState Perl) and install it. Set it PATH so PERL scripts can be executed in every directory.
- 4) Copy merge.pl to the work directory C:\MSMS\work\
If DeconMSn.exe -XMGF is used please adjust all parameters in the BATCH file for MGF file creation including the isolation width and m/z tolerance
- 5) Start mgfprecursor.exe – if an error message occurs install the
Visual C++ 2008 Redistributable Package (x86) vcredist_x86.exe
Microsoft Core XML Services (MSXML) 6.0 msxml6.msi
Microsoft .NET Framework Version 2.0 Redistributable Package (x86) dotnetfx.exe
(For compiling the source make sure to statically link the MFC in the compiler options)
- 6) Download MSCluster (MSCluster.20080609-1.zip) extract the ZIP into the folder
C:\MSMS\MSCluster
Create the directory C:\MSMS\MSCluster\in\
Create the directory C:\MSMS\MSCluster\out\
Create the directory C:\MSMS\MSCluster\tmp\

7) Start the file create-cluster-mfg.bat

That will process the example file PCs_Pos_ID_CE45_01.raw and will create the combined MGF file merge-ok.mgf which will be used in the NIST MS Search GUI.

The MGF file is created in the folder is directly used for library search.

Comment: The mscluster.exe accepts multiple formats dta, mgf, mzXML. Two example batchfiles are provided for msn_extract

Important: The old *.raw file in the folder C:\MSMS\work\ needs to be deleted so only the current *.raw file will be processed. Multiple *.raw files can be processed together.

Congratulations! You created a combined and clustered MGF file with multiple MS/MS scans.

Automated transfer for Thermo XCalibur users

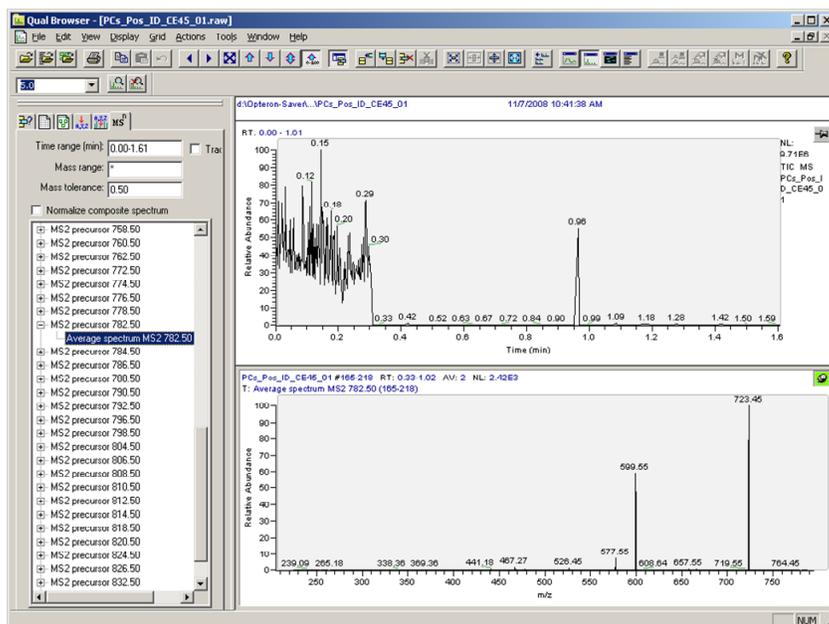
Special section for Thermo XCalibur users
for LC-MS and Infusion MS data

MS/MS spectra are automatically transferred to NIST MS GUI

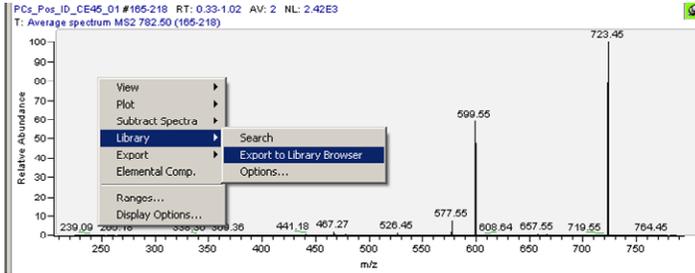
XCalibur currently does not transfer accurate product ion masses, only accurate precursor masses

(that works in principle with every software via the locator
file autoimp.ms and an invoked start of nistms.exe)

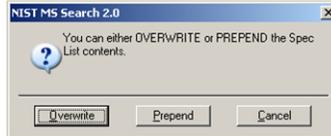
Start XCalibur, goto MS/MS



Right click with mouse in Xcalibur



NIST MS starts automatically

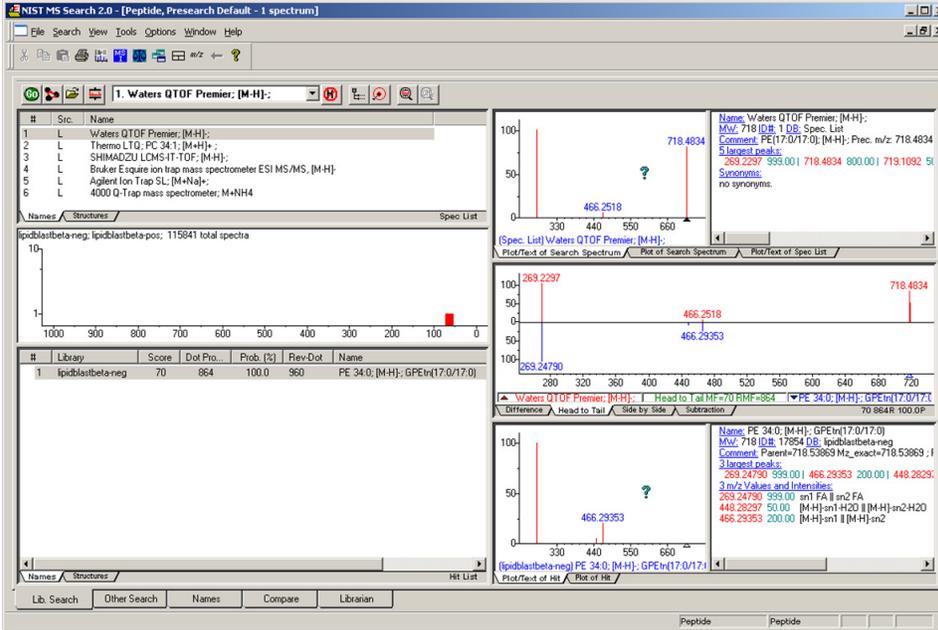


Double click spectrum (1) – positive ID with

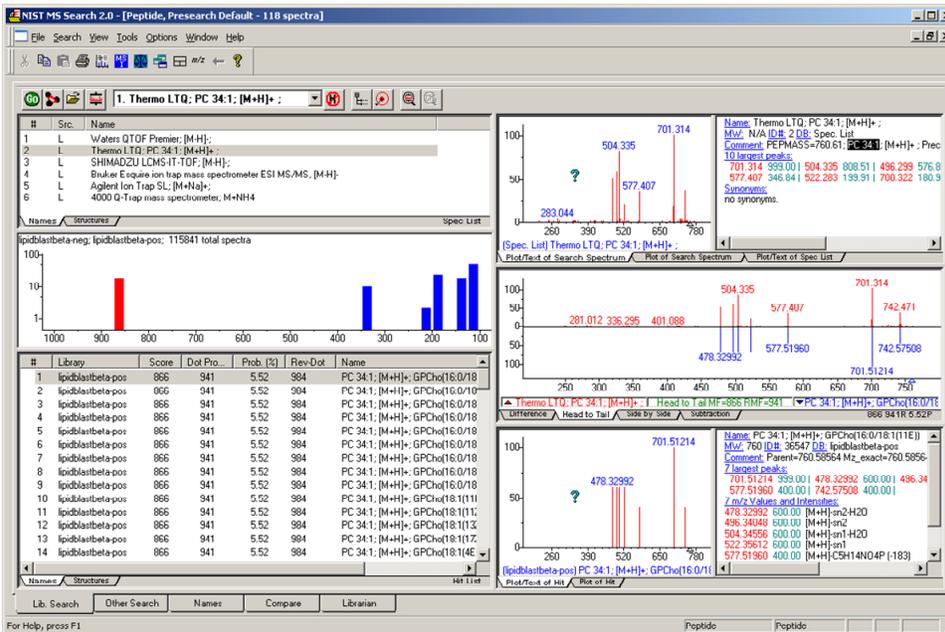
#	Library	Score	Dot Pro.	Prob. (%)	Rev-Dot	Name
1	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
2	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
3	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
4	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
5	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
6	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
7	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
8	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
9	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(16:0)1
10	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(18:1)1
11	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(18:1)1
12	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(18:1)1
13	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(18:1)1
14	lipidblastbeta-pos	554	949	5.34	979	PC 34:1; [M+Na]+; GPChol(18:1)4

Single search examples for ABI, Agilent, Bruker, Thermo, Waters

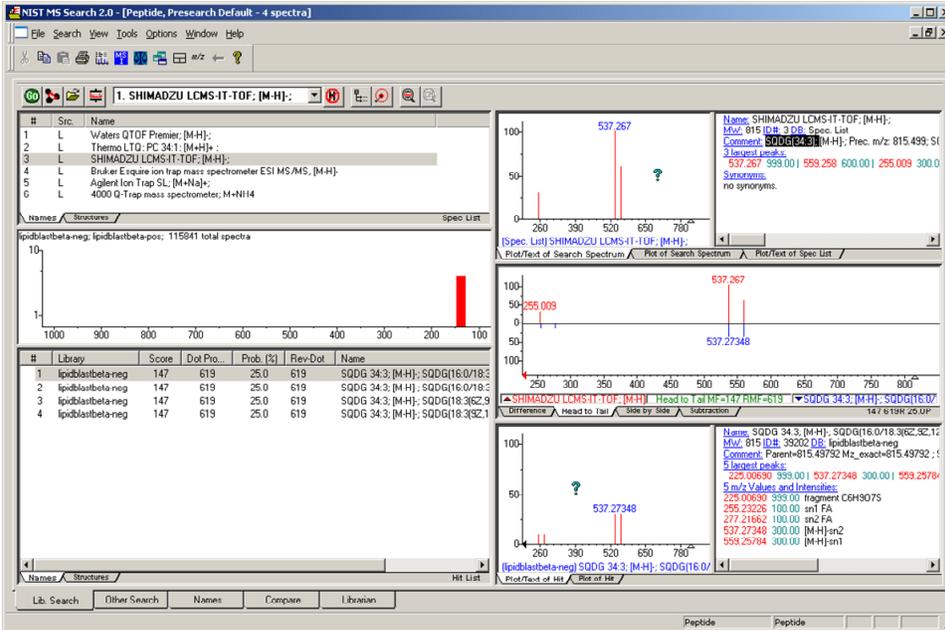
Example 1) Name: Waters QTOF Premier; [M-H]⁻; PE(17:0/17:0);



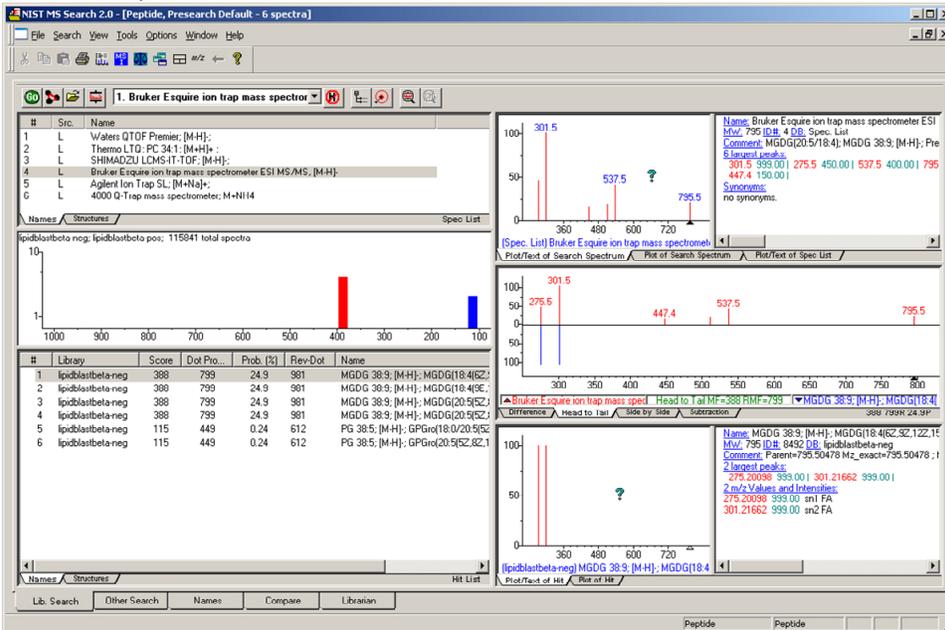
Example 2) Name: Thermo LTQ; PC 34:1; [M+H]⁺; PC 34:1



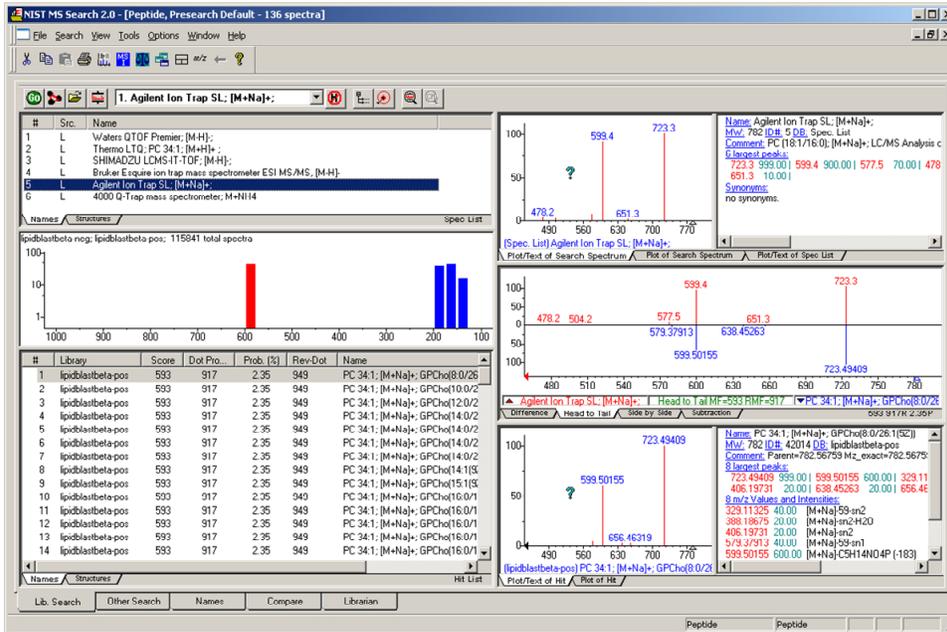
Example 3) Name: SHIMADZU LCMS-IT-TOF; [M-H]⁻; SQDG(34:3);



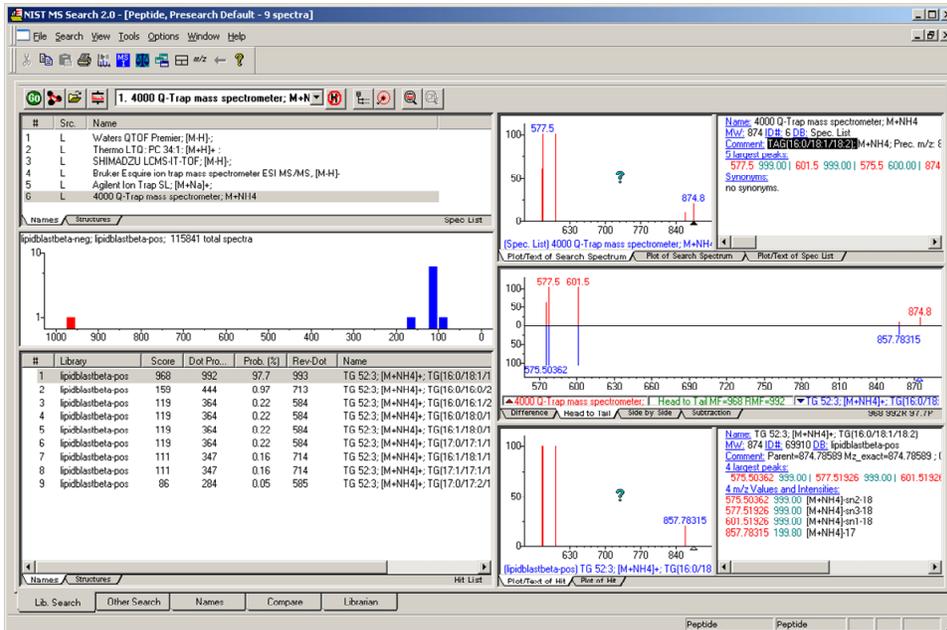
Example 4) Name: Bruker Esquire ion trap mass spectrometer ESI MS/MS, [M-H]⁻; MGDG 38:9;



Example 5) Name: Agilent Ion Trap SL; [M+Na]⁺; PC (18:1/16:0);



Example 6) ABI Name: 4000 Q-Trap mass spectrometer; M+NH₄⁺; TAG(16:0/18:1/18:2);



How does the text based format look like?

The library is natively stored in NIST MSP text format. That allows an easy conversion to all possible machine specific or program specific formats.

Name: PC 36:4; [M+H]⁺; GPCho(20:4(7E,10E,13E,16E)/16:0)

MW: 782.56995

PRECURSORMZ: 782.56995

Comment: Parent=782.56995 Mz_exact=782.56995 ; PC 36:4; [M+H]⁺; GPCho(20:4(7E,10E,13E,16E)/16:0); C44H80NO8P

Num Peaks: 7

764.55939 400 "[M+H]-H2O (-18)"

723.49645 999 "[M+H]-C3H9N (-59)"

599.50391 400 "[M+H]-C5H14NO4P (-183)"

544.34043 600 "[M+H]-sn2"

526.32987 600 "[M+H]-sn2-H2O"

496.34043 600 "[M+H]-sn1"

478.32987 600 "[M+H]-sn1-H2O"

Due to the standardized format the library is versatile and universal applicable. Fragments are annotated as accurate masses and specific fragments are named.

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

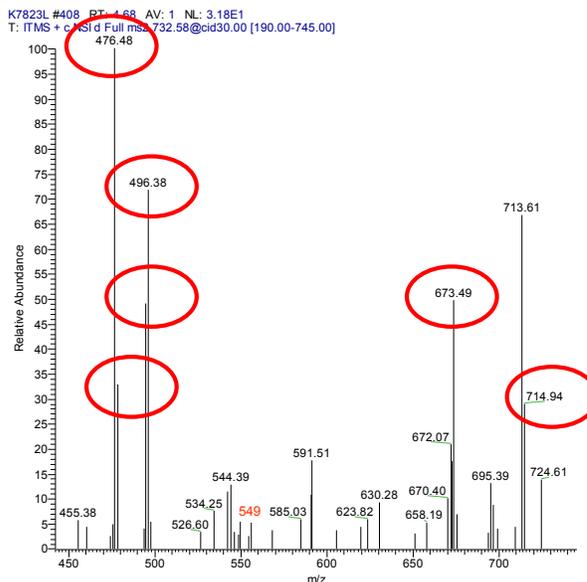


Figure: The in-silico fragment (left) match the experimental MS/MS spectra (right). The peak abundances in experimental MS/MS spectra may fluctuate or in case of overlapping compounds may differ in some cases from the in-silico model.

I want to integrate LipidBlast with my own software program

The LipidBlast libraries available under the Creative Commons by attribution (CC-BY) license.

LipidBlast - In silico created MS/MS libraries for lipid profiling
License (CC-by) Tobias Kind; Metabolomics Fiehn Lab (2009)
<http://fiehnlab.ucdavis.edu/staff/kind/>

<http://creativecommons.org/licenses/by/3.0/>

CC-BY means you are free:

to Share - to copy, distribute and transmit the work

to Remix - to adapt the work

Under the following conditions:

Attribution - You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

There are three easy options:

1) Direct Integration.

Load the MSP files LipidBlastBeta-neg.msp and LipidBlastBeta-pos.msp via an internal parser into your vendor specific mass spectral search program. Make sure to implement an accurate precursor window search and an additional MS/MS matching (dot product). Add the CC-BY license.

2) Automatic transfer to NIST MS Search GUI

Your program transfers the mass spectral file into NIST MS Search GUI. It uses and locator file (autoimp.ms). The accurate mass file with precursor and product ions is transferred to the NIST MS Search GUI and a function call invokes the automatic start of the NIST MS Search GUI. Please refer to <http://chemdata.nist.gov/> for DLL integration and all technical question.

3) Use of MGF (compiled *.dta) file

This method is used by Hundreds of proteomics groups and supported by all major mass spectrometry vendors. NIST MS GUI can read *.MGF files. Please use Google to find out more.

<http://www.google.com/#num=100&hl=en&safe=active&q=MGF+and+DTA+files>