www.lipidmaps.org

Tools for lipidomics: MS prediction and structure manipulation methods

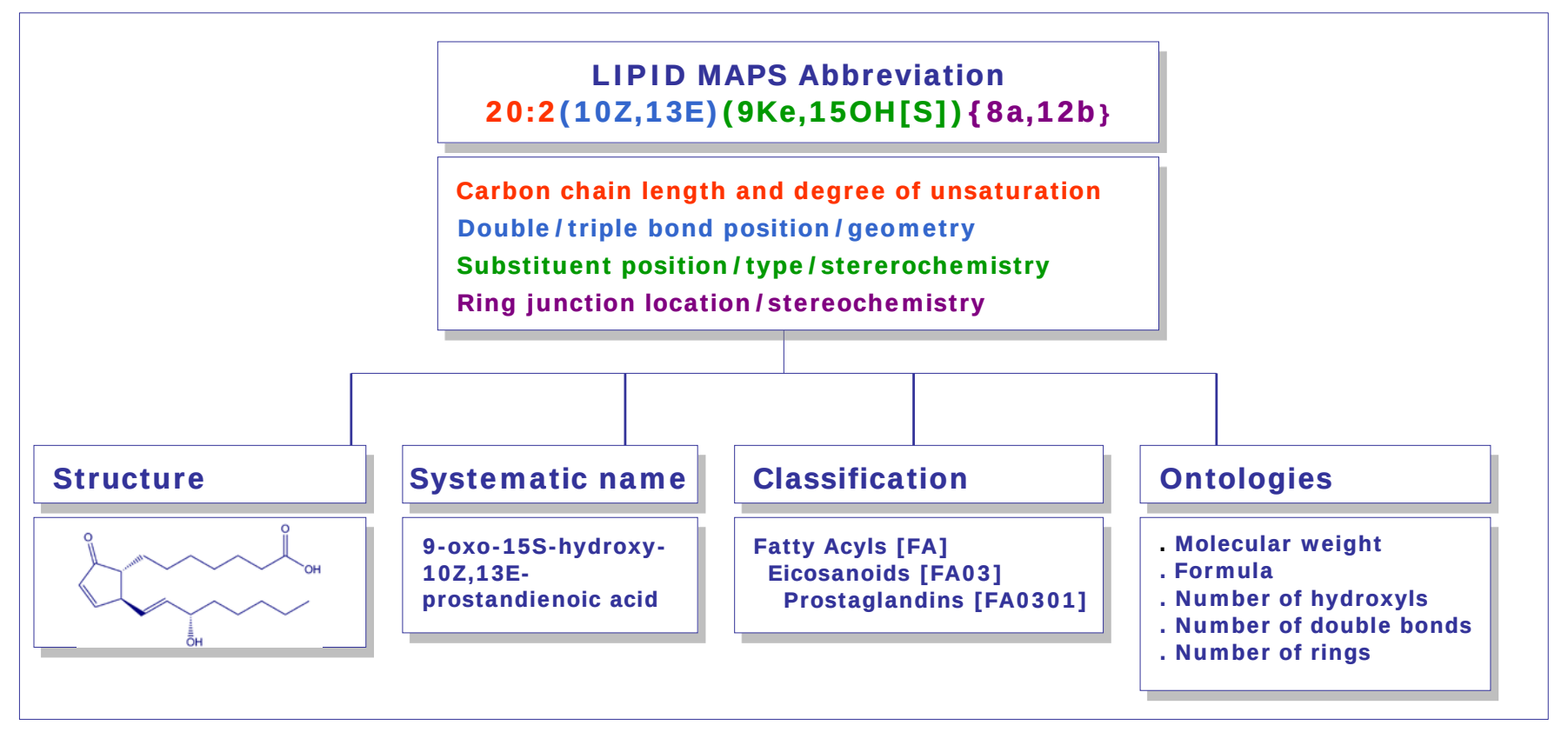
Dawn Cotter[a], Eoin Fahy[a], Manish Sud[a], Shankar Subramaniam[a,b]

National Institute of General Medical Sciences
LIPID MAPS funded by Glue
Grant from :www.nigms.nih.gov

Introduction

Certain classes of lipids such as glycerolipids, glycerophospholipids and sphingolipids which are composed of an invariant core and one or more acyl/alkyl substituents are good candidates for MS computational analysis since their molecular ion masses can usually be predicted on the basis of permutations of commonly occurring side-chains and also because they tend to fragment predictably in EIMS leading to detection of diagnostic product ions. It is possible to create a virtual database of permutations of the more common side-chains for various lipid categories and calculate "high-probability" product ion candidates in order to compare experimental data with predicted spectra. The LIPID MAPS group has developed a suite of search tools allowing a user to enter an m/z value of interest and view a list of matching structure candidates, along with a list of calculated "high-probability" product ions. These search interfaces have been integrated with structure-drawing and isotopic-distribution tools. A suite of structure-drawing tools has been developed and deployed which dramatically increase the efficiency of generating large and complex lipid structures, facilitate data entry into lipid-structure databases and permit "on-demand" structure generation in conjunction with a variety of mass spectrometry–based informatics tools.

Lipid structure manipulation and ontology tools

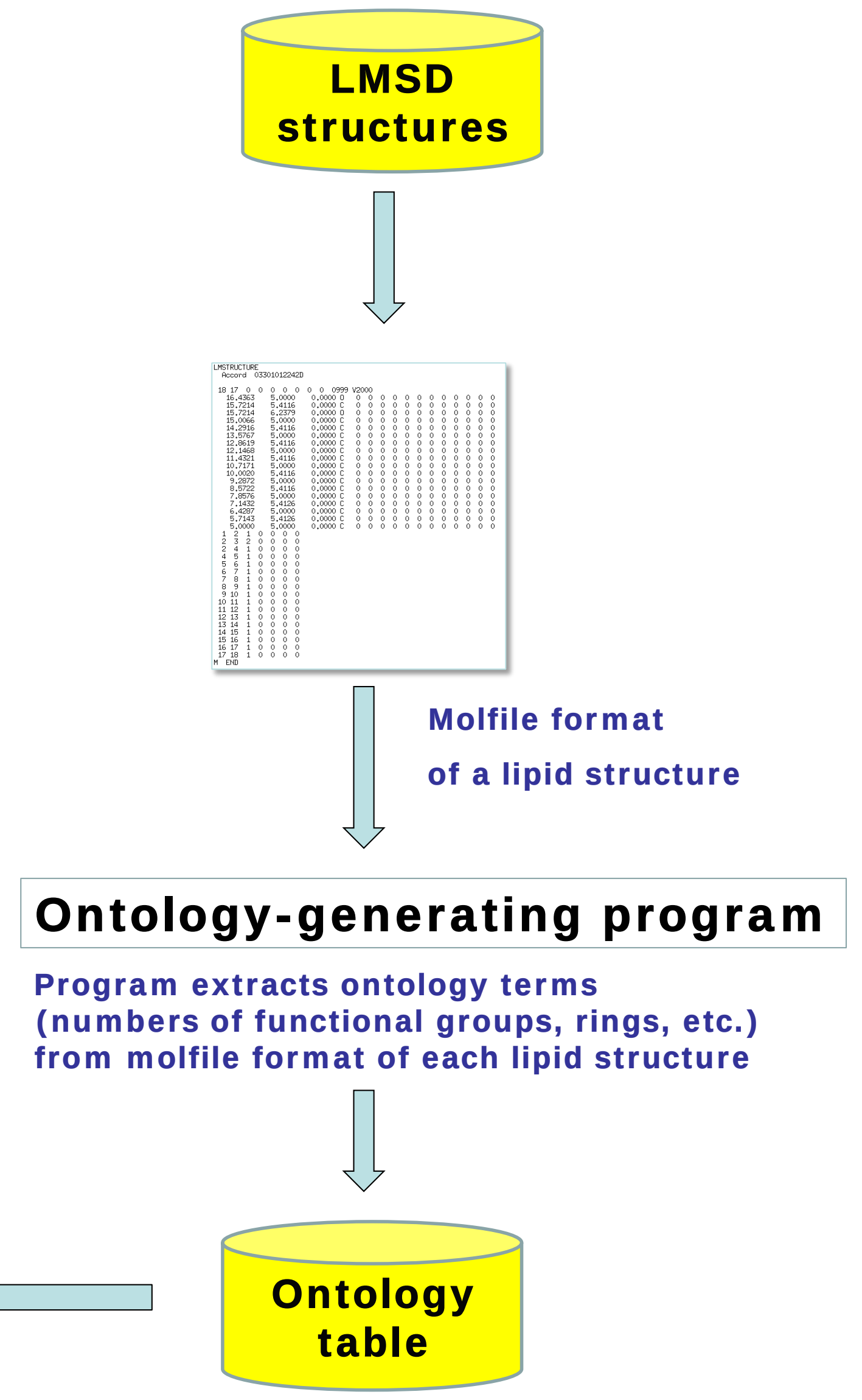


LIPID MAPS Abbreviation
20:2(10Z,13E)(9Ke,15OH[S]){8a,12b}

Carbon chain length and degree of unsaturation
Double/triple bond position/geometry
Substituent position/type/stererochemistry
Ring junction location/stereochemistry

Structure	Systematic name	Classification	Ontologies
	9-oxo-15S-hydroxy-10Z,13E-prostandienoic acid	Fatty Acyls [FA] Eicosanoids [FA03] Prostaglandins [FA0301]	. Molecular weight . Formula . Number of hydroxyls . Number of double bonds . Number of rings

Generation of structure and ontology from abbreviation



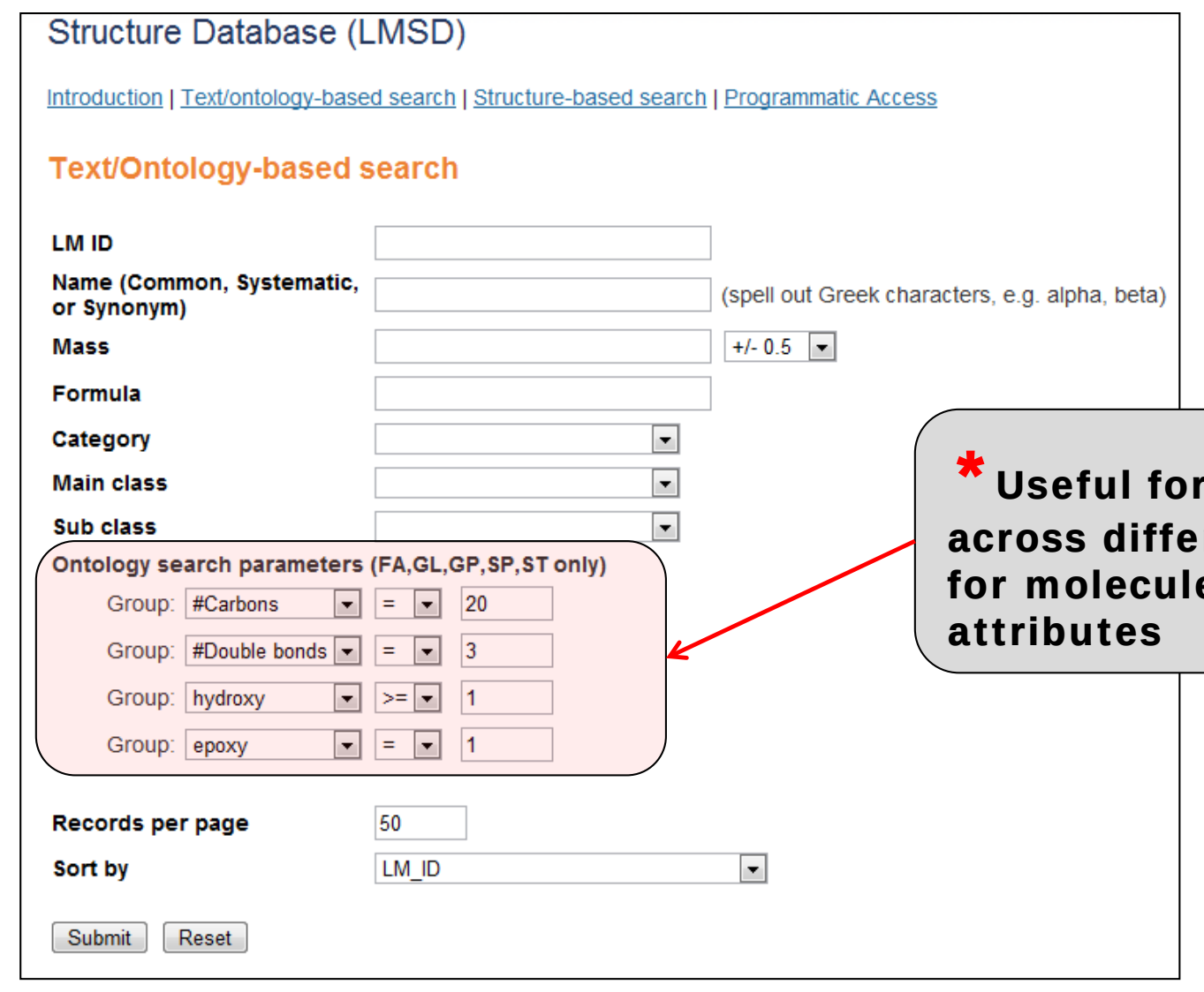
LMSD structures

Molfile format of a lipid structure

Ontology-generating program

Program extracts ontology terms (numbers of functional groups, rings, etc.) from molfile format of each lipid structure

Ontology table



Structure Database (LMSD)

Introduction | Text/ontology-based search | Structure-based search | Programmatic Access

Text/Ontology-based search

LM ID
Name (Common, Systematic, or Synonym) (spell out Greek characters, e.g. alpha, beta)
Mass: +/- 0.5
Formula
Category
Main class
Sub class

Ontology search parameters (FA,GL,GP,SP,ST only)
Group: #Carbons = 20
Group: #Double bonds = 3
Group: hydroxy >= 1
Group: epoxy = 1

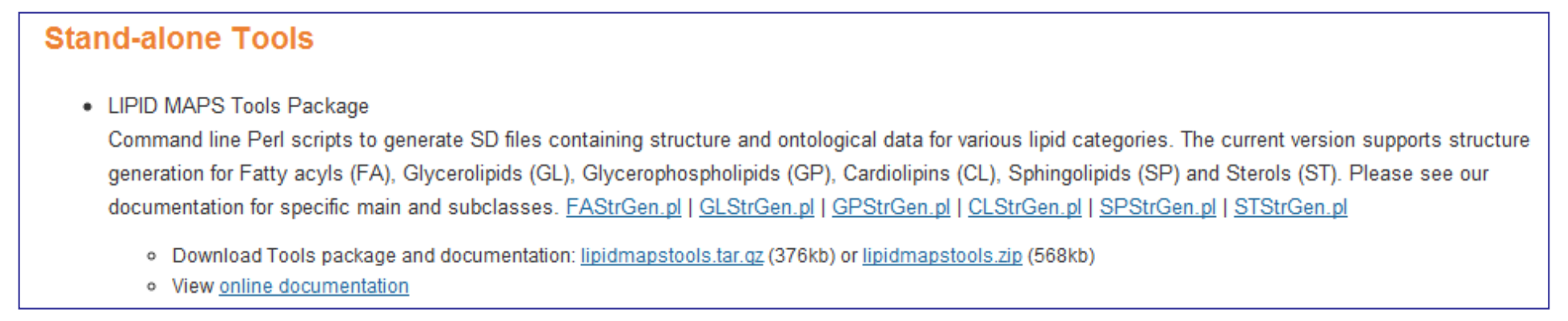
Records per page: 50
Sort by: LM_ID

* Useful for searching across different lipid classes for molecules with certain attributes

Search LMSD with ontology terms

Standalone and command-line tools

Perl scripts for lipid structure/name/mass generation

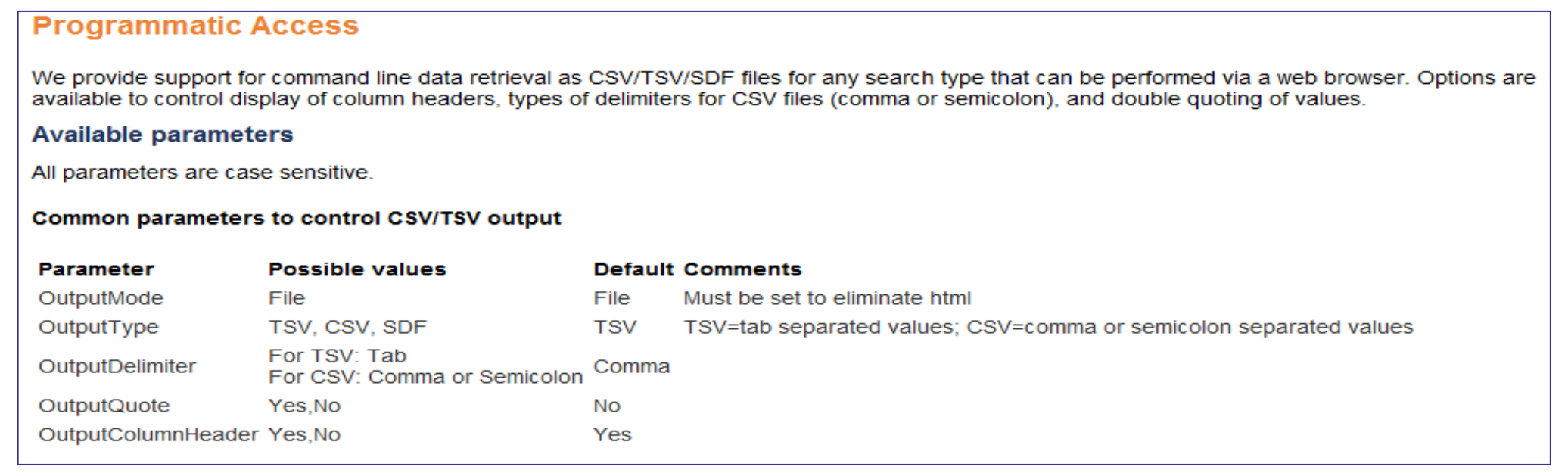


Stand-alone Tools

- LIPID MAPS Tools Package
  Command line Perl scripts to generate SD files containing structure and ontological data for various lipid categories. The current version supports structure generation for Fatty acyls (FA), Glycerolipids (GL), Glycerophospholipids (GP), Cardiolipins (CL), Sphingolipids (SP) and Sterols (ST). Please see our documentation for specific main and subclasses. FAStrGen.pl | GLStrGen.pl | GPStrGen.pl | CLStrGen.pl | SPStrGen.pl | STStrGen.pl
  - Download Tools package and documentation: lipidmapstools.tar.gz (376kb) or lipidmapstools.zip (568kb)
  - View online documentation

Use a "wget" command to retrieve lipid structures and/or tabular attributes

http::.www.lipidmaps.org/data/structure/programmaticaccess.html



Programmatic Access

We provide support for command line data retrieval as CSV/TSV/SDF files for any search type that can be performed via a web browser. Options are available to control display of column headers, types of delimiters for CSV files (comma or semicolon), and double quoting of values.

Available parameters

All parameters are case sensitive.

Common parameters to control CSV/TSV output

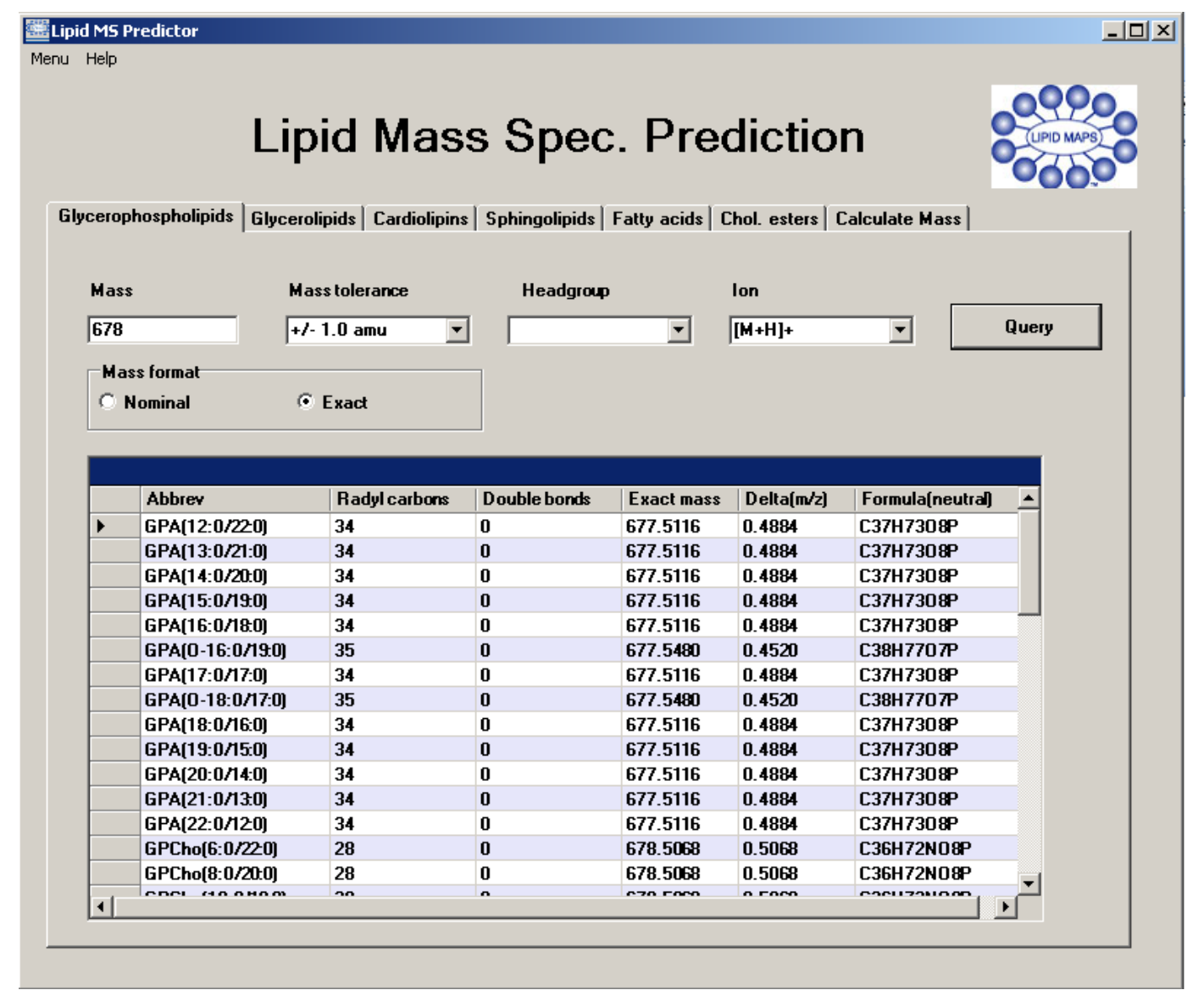
Parameter	Possible values	Default	Comments
OutputMode	File	File	Must be set to eliminate html
OutputType	TSV, CSV, SDF	TSV	TSV=tab separated values; CSV=comma or semicolon separated values
OutputDelimiter	For TSV: Tab For CSV: Comma or Semicolon	Comma	
OutputQuote	Yes,No	No	
OutputColumnHeader	Yes,No	Yes	

Example: Find all Diacylglycerols [GL0201] with mass between 584.00 and 604.00 daltons

```
wget -O "Test" "http://www.lipidmaps.org/data/structure/LMSDSearch.php?Mode=ProcessTextSearch\
&OutputMode=File&CoreClass=2&MainClass=202&SubClass=20201&ExactMass=594&ExactMassOffSet=
10"
```

LIPID MAPS MS Prediction Tool (Windows application)

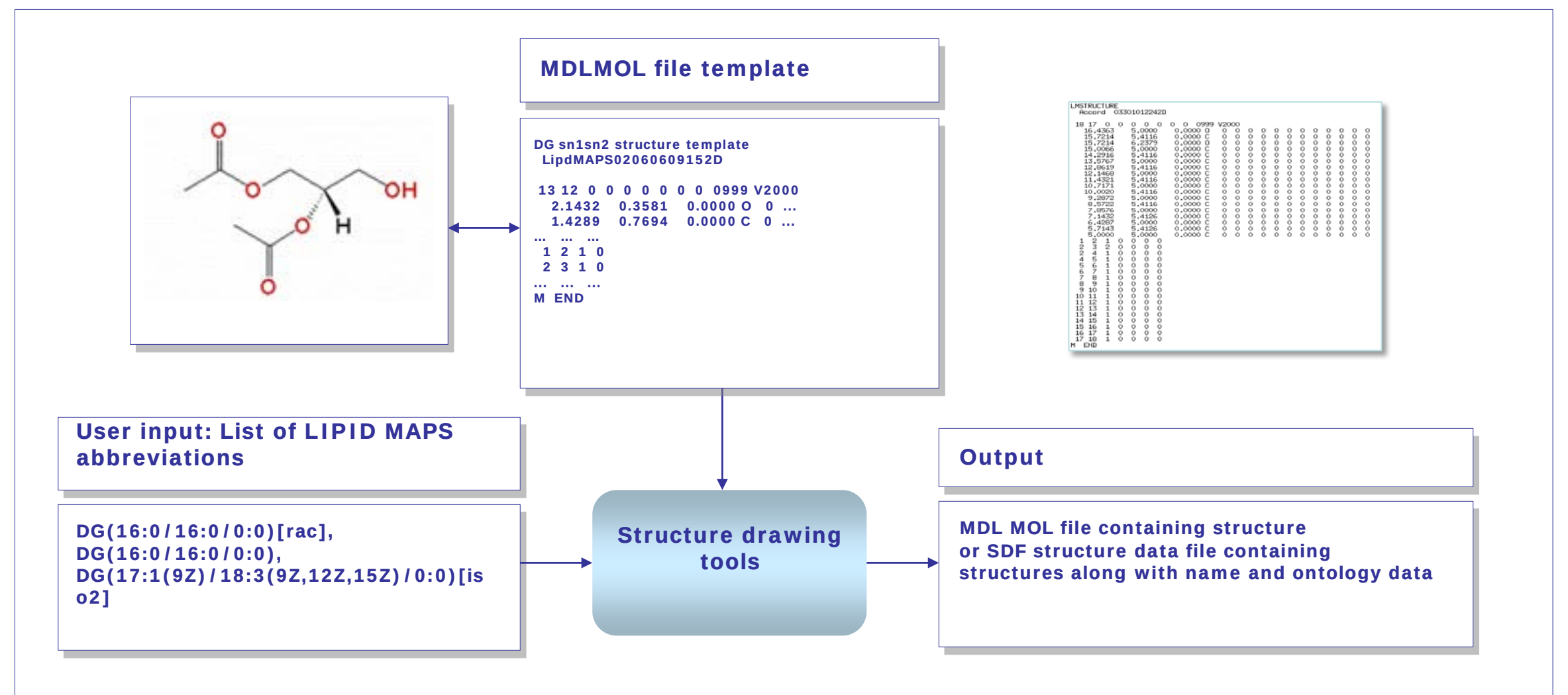
Search interfaces for glycerolipids, glycerophospholipids, cardiolipins, sphingolipids, fatty acids and cholesteryl esters with user-supplied m/z values



Mass: 678
Mass tolerance: +/- 1.0 amu
Headgroup:
Ion: [M+H]+
Mass format: ○ Nominal ● Exact

Abbrev	Radyl carbons	Double bonds	Exact mass	Delta(m/z)	Formula(neutral)
GPA(12:0/22:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(13:0/21:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(14:0/20:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(15:0/19:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(16:0/18:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(O-16:0/19:0)	35	0	677.5480	0.4520	C38H77O7P
GPA(17:0/17:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(O-18:0/17:0)	35	0	677.5480	0.4520	C38H77O7P
GPA(18:0/16:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(19:0/15:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(20:0/14:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(21:0/13:0)	34	0	677.5116	0.4884	C37H73O8P
GPA(22:0/12:0)	34	0	677.5116	0.4884	C37H73O8P
GPCho(6:0/22:0)	28	0	678.5068	0.5068	C36H72NO8P
GPCho(8:0/20:0)	28	0	678.5068	0.5068	C36H72NO8P

Structure drawing tools



MDLMOL file template

```
DG sn1sn2 structure template
 LipdMAPS02060609152D

 13 12  0  0  0  0  0  0  0999 V2000
   2.1432   0.3581   0.0000 O   0 ...
   1.4289   0.7694   0.0000 C   0 ...
 ...  ...  ...
   1  2  1  0
   2  3  1  0
 ...  ...  ...
M  END
```

User input: List of LIPID MAPS abbreviations

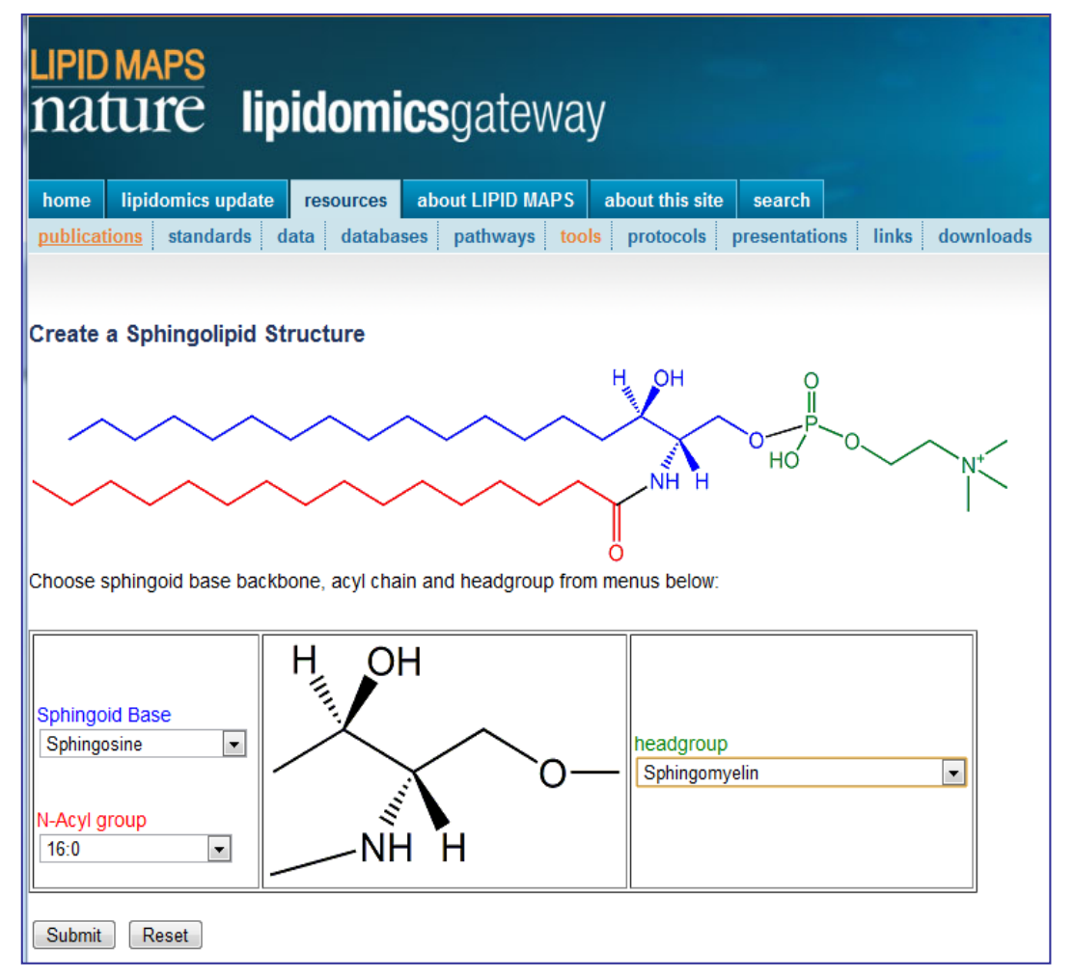
DG(16:0/16:0/0:0)[rac], DG(16:0/16:0/0:0), DG(17:1(9Z)/18:3(9Z,12Z,15Z)/0:0)[iso2]

Structure drawing tools

Output

MDL MOL file containing structure or SDF structure data file containing structures along with name and ontology data

Structure generation methodology

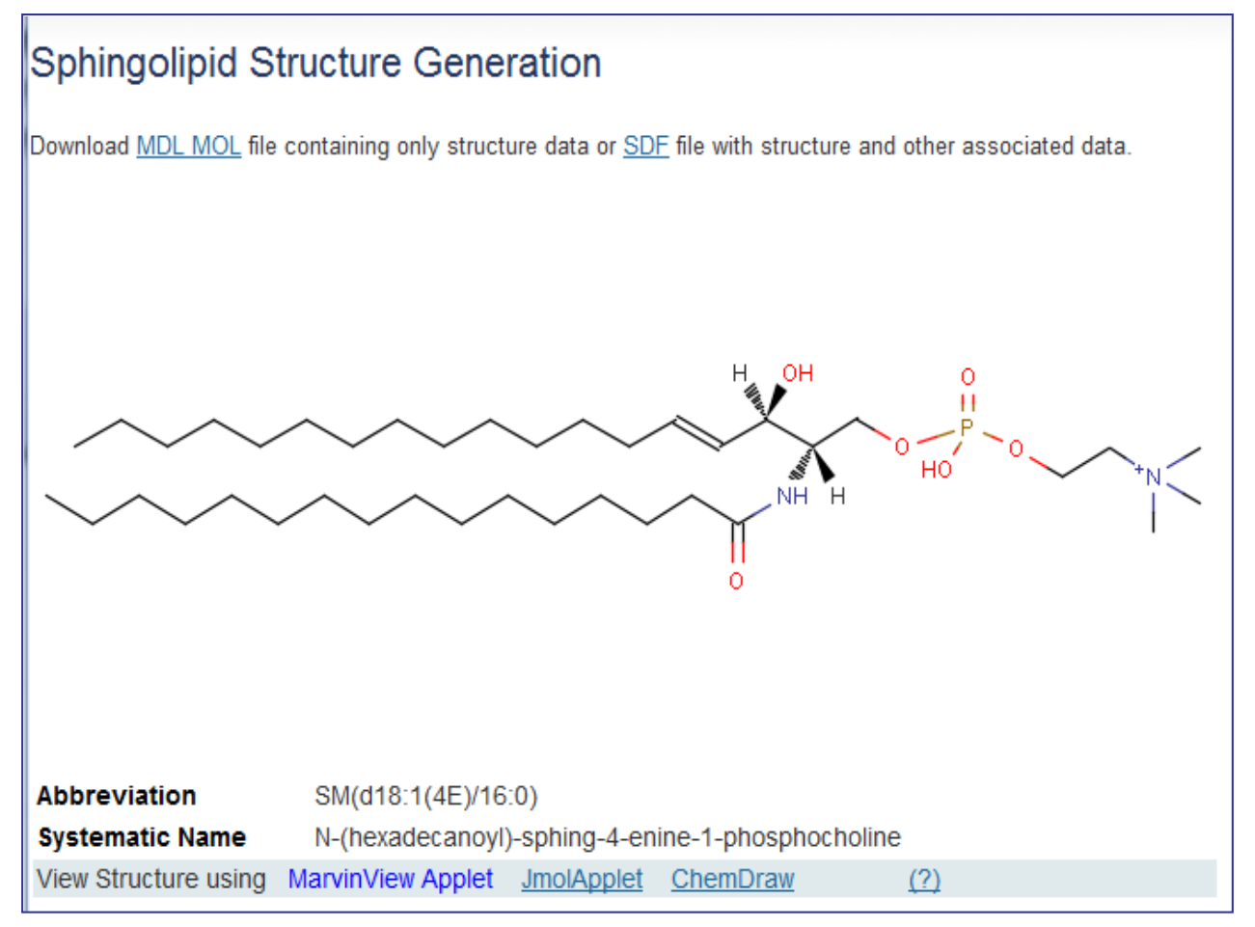


Create a Sphingolipid Structure

Choose sphingoid base backbone, acyl chain and headgroup from menus below:

Sphingoid Base: Sphingosine
N-Acyl group: 16:0
headgroup: Sphingomyelin

Online drawing tools for various lipid categories (FA,GL,GP,SP,ST)



Sphingolipid Structure Generation

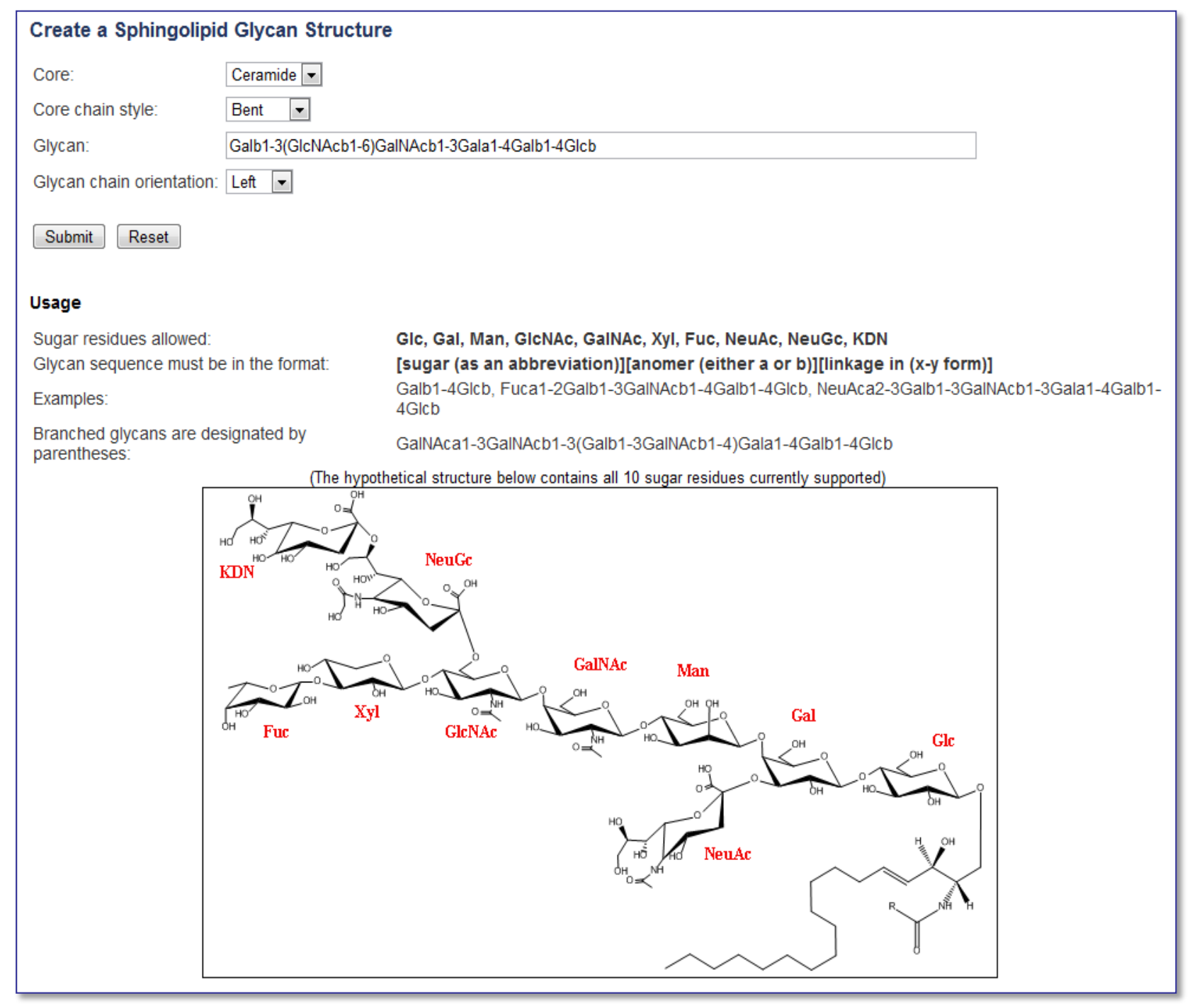
Download MDL MOL file containing only structure data or SDF file with structure and other associated data.

Abbreviation: SM(d18:1(4E)/16:0)
Systematic Name: N-(hexadecanoyl)-sphing-4-enine-1-phosphocholine
View Structure using MarvinView Applet JmolApplet ChemDraw (?)

Structures viewable in Marvin, JMol and Chemdraw format
May be saved as Molfiles

Sugars
Glc
Gal
GlcNAc
GalNac
Xyl
Fuc
Man
NeuAc
NeuGc
KDN

Anomeric Carbon
α or β linkages may be specified



Create a Sphingolipid Glycan Structure

Core: Ceramide
Core chain style: Bent
Glycan: Galb1-3(GlcNAcb1-6)GalNAcb1-3Gala1-4Galb1-4Glcb
Glycan chain orientation: Left

Usage

Sugar residues allowed:	Glc, Gal, Man, GlcNAc, GalNAc, Xyl, Fuc, NeuAc, NeuGc, KDN
Glycan sequence must be in the format:	[sugar (as an abbreviation)][anomer (either a or b)][linkage in (x-y form)]
Examples:	Galb1-4Glcb, Fuca1-2Galb1-3GalNAcb1-4Galb1-4Glcb, NeuAca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb
Branched glycans are designated by parentheses:	GalNAca1-3GalNAcb1-3(Galb1-3GalNAcb1-4)Gala1-4Galb1-4Glcb

(The hypothetical structure below contains all 10 sugar residues currently supported)

All lipid categories have examples of glycosylated structures

Generation of glycan structures in full chair conformation

Example:Pherobase fatty acids

C14

(E)-5-Tetradecenoic acid	226.36	C14H26O2
(E)-7-oxo-11,13-Tetradecadienoic acid	238.32	C14H22O3
(E)-7-oxo-11-Tetradecenoic acid	240.34	C14H24O3
(E,Z)-3,5-Tetradecadienoic acid	224.34	C14H24O2
(Z)-5-Tetradecenoic acid	226.36	C14H26O2
(Z)-7-oxo-11-Tetradecenoic acid	240.34	C14H24O3
(Z,Z)-3,5-Tetradecadienoic acid	224.34	C14H24O2
3-Hydroxytetradecanoic acid	244.37	C14H28O3
5,8-Tetradecadienoic acid	224.34	C14H24O2
7-Hydroxytetradecanoic acid	244.37	C14H28O3
7-Oxotetradecanoic acid	242.35	C14H26O3
7-Tetradecenoic acid	226.36	C14H26O2
Tetradecadienoic acid	224.34	C14H24O2
Tetradecanoic acid	228.37	C14H28O2
Tetradecenoic acid	226.36	C14H26O2

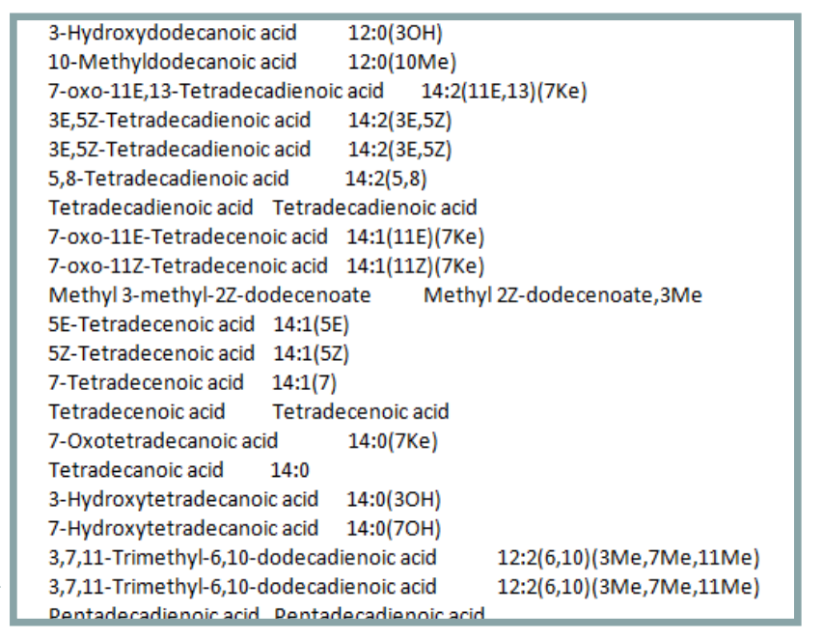
C15

(2S,7S)-2,7-Nonanediyldibutyrate	300.43	C17H32O4
O-(Z)-8-Pentadecenylsalicylic acid	346.5	C22H34O3
Pentadecadienoic acid	238.37	C15H26O2
Pentadecanoic acid	242.4	C15H30O2

C16

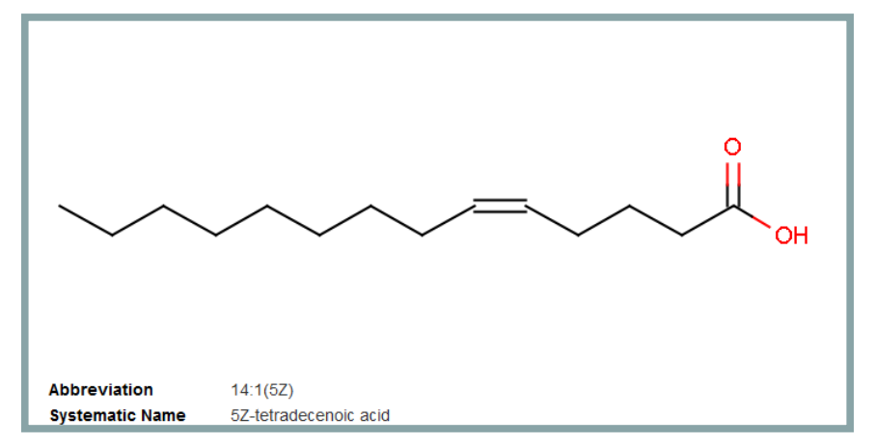
(Z)-11-Hexadecenoic acid	254.41	C16H30O2
(Z)-9-Hexadecenoic acid	254.41	C16H30O2

List of systematic names



3-Hydroxydodecanoic acid	12:0(3OH)
10-Methyldodecanoic acid	12:0(10Me)
7-oxo-11E,13-Tetradecadienoic acid	14:2(11E,13)(7Ke)
3E,5Z-Tetradecadienoic acid	14:2(3E,5Z)
3E,5Z-Tetradecadienoic acid	14:2(3E,5Z)
5,8-Tetradecadienoic acid	14:2(5,8)
Tetradecadienoic acid	Tetradecadienoic acid
7-oxo-11E-Tetradecenoic acid	14:1(11E)(7Ke)
7-oxo-11Z-Tetradecenoic acid	14:1(11Z)(7Ke)
Methyl 3-methyl-2Z-dodecenoate	Methyl 2Z-dodecenoate,3Me
5E-Tetradecenoic acid	14:1(5E)
5Z-Tetradecenoic acid	14:1(5Z)
7-Tetradecenoic acid	14:1(7)
Tetradecenoic acid	Tetradecenoic acid
7-Oxotetradecanoic acid	14:0(7Ke)
Tetradecanoic acid	14:0
3-Hydroxytetradecanoic acid	14:0(3OH)
7-Hydroxytetradecanoic acid	14:0(7OH)
3,7,11-Trimethyl-6,10-dodecadienoic acid	12:2(6,10)(3Me,7Me,11Me)
3,7,11-Trimethyl-6,10-dodecadienoic acid	12:2(6,10)(3Me,7Me,11Me)

List of abbreviations



Abbreviation: 14:1(5Z)
Systematic Name: 5Z-tetradecenoic acid

Structures (SD file)

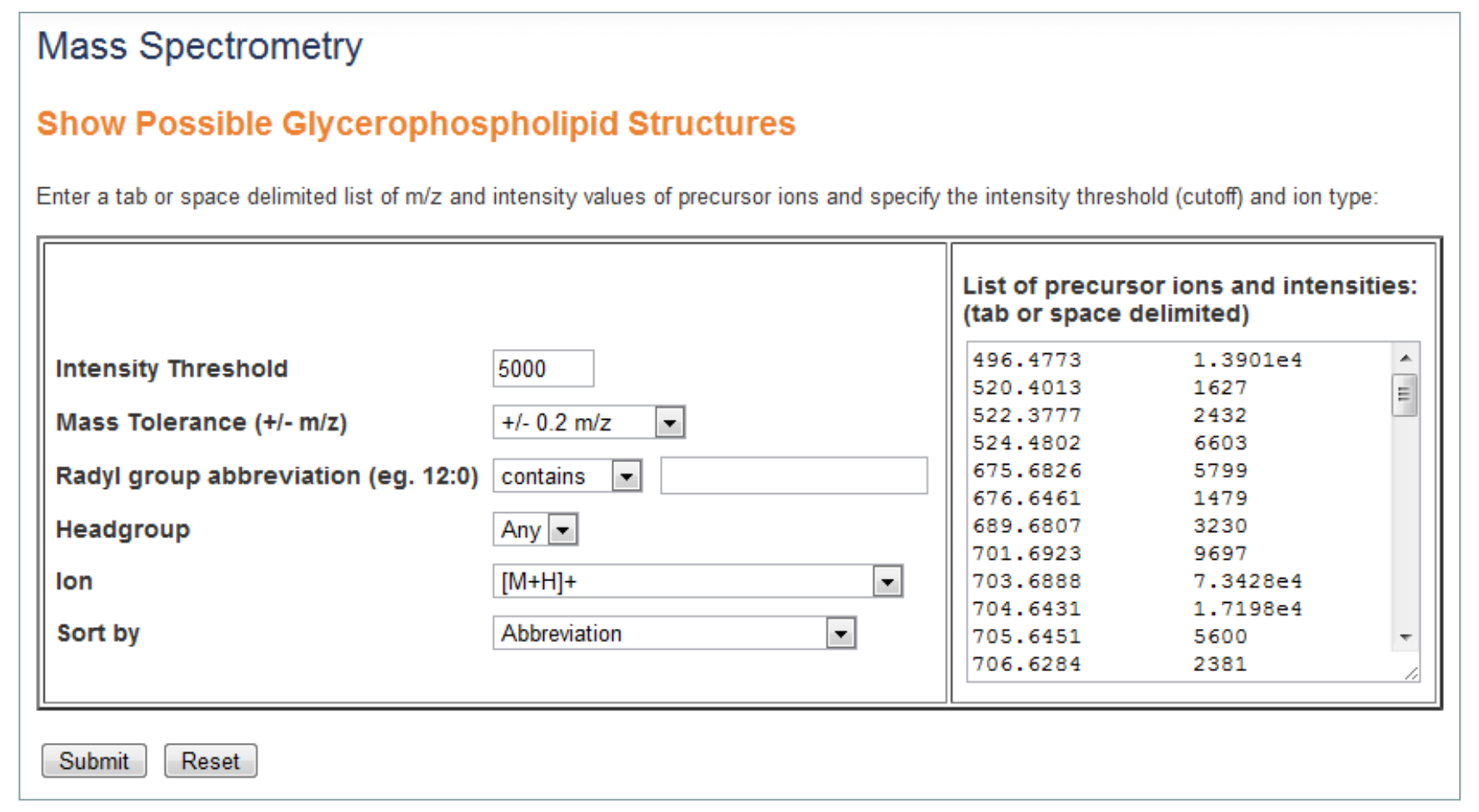
Computational generation of structures from names/abbreviations

Mass spectrometry prediction tools[1]

[1] LIPID MAPS online tools for lipid research. Nucleic Acid Research, 2007, Vol. , Web server issue W606-612

MS prediction using virtual databases of structures based on commonly occurring core structures and chains

Data entry (precursor ion list, search criteria)



Mass Spectrometry

Show Possible Glycerophospholipid Structures

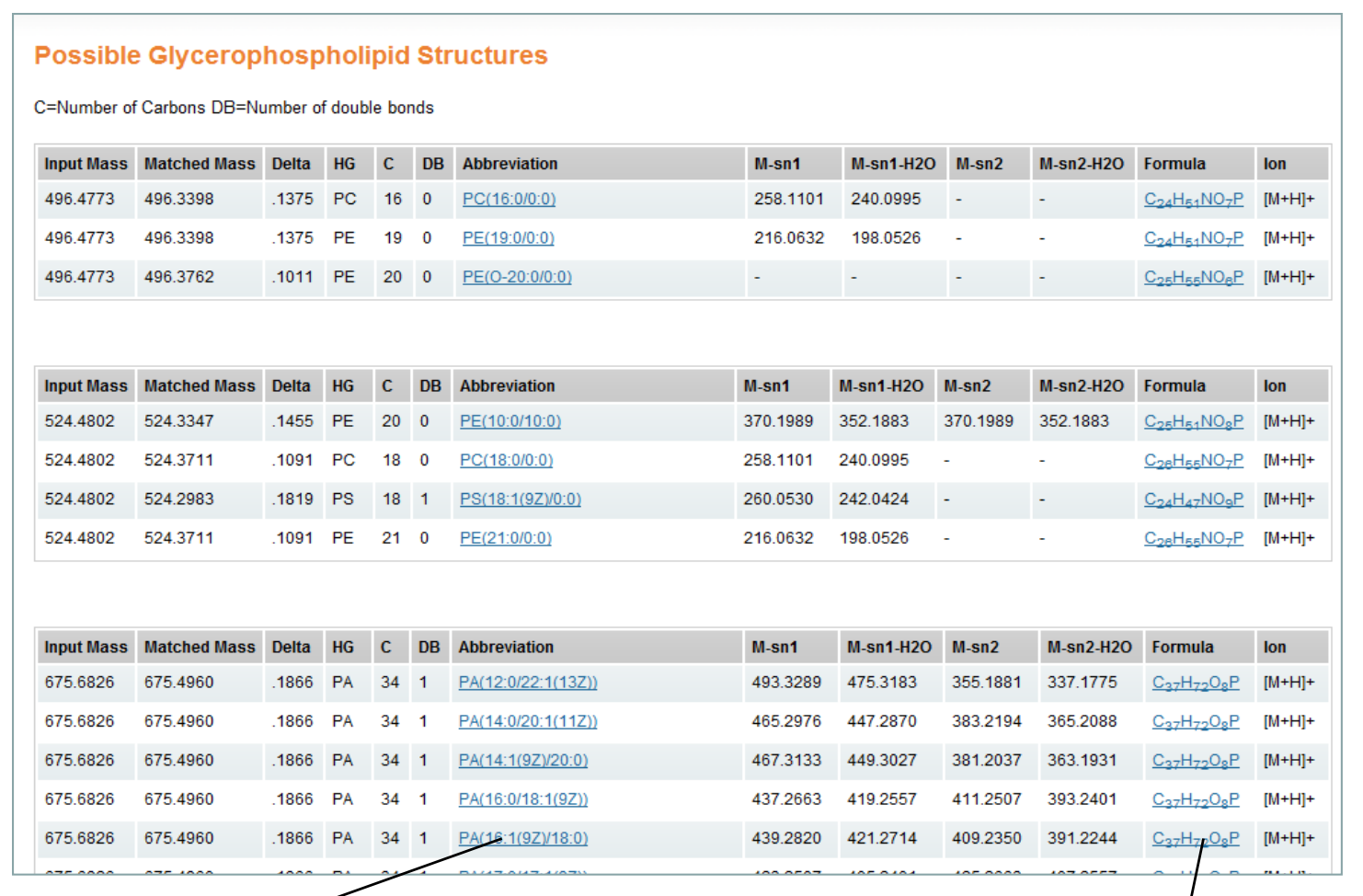
Enter a tab or space delimited list of m/z and intensity values of precursor ions and specify the intensity threshold (cutoff) and ion type:

Intensity Threshold: 5000
Mass Tolerance (+/- m/z): +/- 0.2 m/z
Radyl group abbreviation (eg. 12:0): contains
Headgroup: Any
Ion: [M+H]+
Sort by: Abbreviation

List of precursor ions and intensities: (tab or space delimited)

```
496.4773      1.3901e4
520.4013      1627
522.3777      2432
524.4802      6603
675.6826      5799
676.6461      1479
689.6807      3230
701.6923      9697
703.6888      7.3428e4
704.6431      1.7198e4
705.6451      5600
706.6284      2381
```

Search virtual database of GP masses



Possible Glycerophospholipid Structures

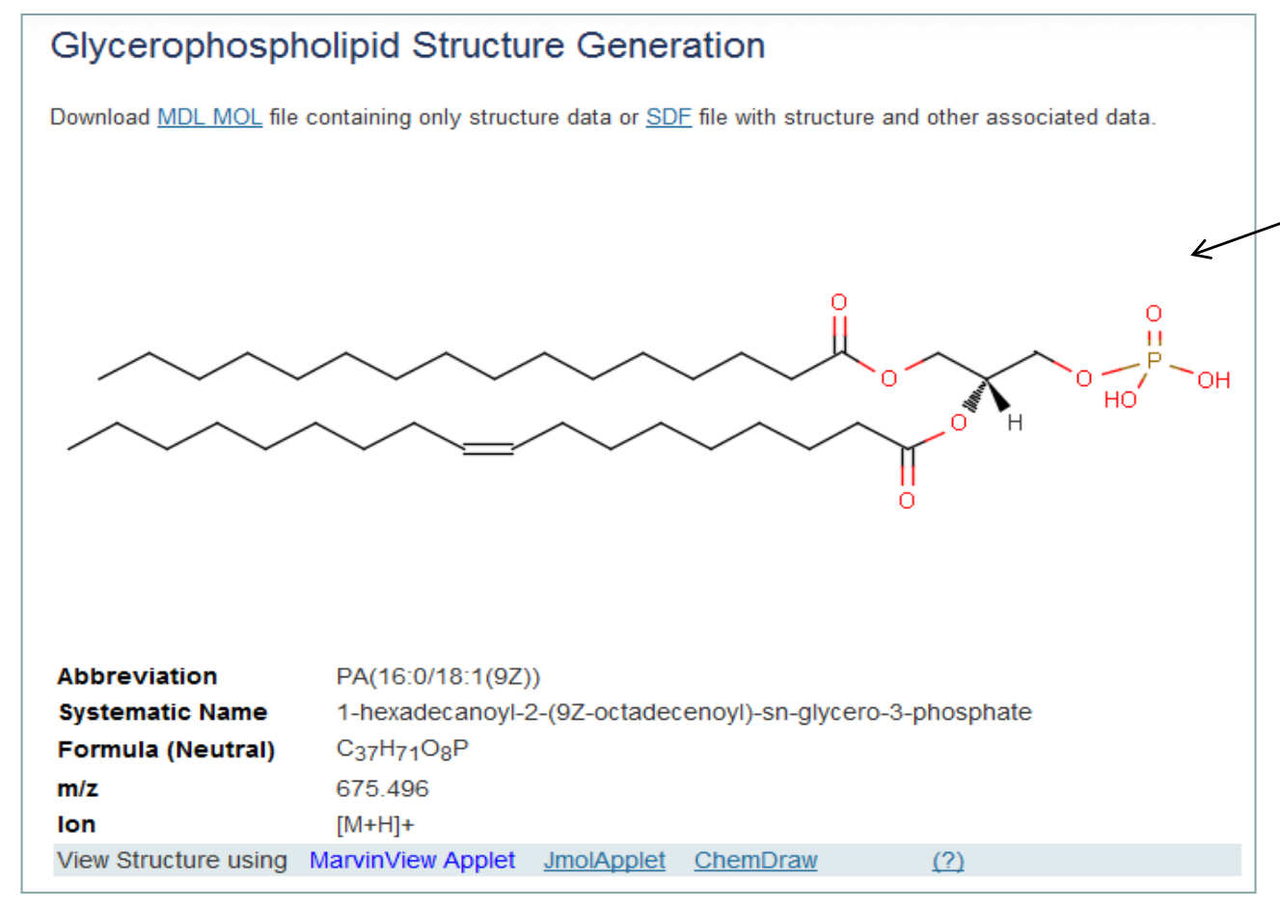
C=Number of Carbons DB=Number of double bonds

Input Mass	Matched Mass	Delta	HG	C	DB	Abbreviation	M-sn1	M-sn1-H2O	M-sn2	M-sn2-H2O	Formula	Ion
496.4773	496.3398	.1375	PC	16	0	PC(16:0/0:0)	258.1101	240.0995	-	-	$C_{24}H_{51}NO_7P$	[M+H]+
496.4773	496.3398	.1375	PE	19	0	PE(19:0/0:0)	216.0632	198.0526	-	-	$C_{24}H_{51}NO_7P$	[M+H]+
496.4773	496.3762	.1011	PE	20	0	PE(O-20:0/0:0)	-	-	-	-	$C_{25}H_{55}NO_6P$	[M+H]+

Input Mass	Matched Mass	Delta	HG	C	DB	Abbreviation	M-sn1	M-sn1-H2O	M-sn2	M-sn2-H2O	Formula	Ion
524.4802	524.3347	.1455	PE	20	0	PE(10:0/10:0)	370.1989	352.1883	370.1989	352.1883	$C_{25}H_{51}NO_8P$	[M+H]+
524.4802	524.3711	.1091	PC	18	0	PC(18:0/0:0)	258.1101	240.0995	-	-	$C_{26}H_{55}NO_7P$	[M+H]+
524.4802	524.2983	.1819	PS	18	1	PS(18:1(9Z)/0:0)	260.0530	242.0424	-	-	$C_{24}H_{47}NO_9P$	[M+H]+
524.4802	524.3711	.1091	PE	21	0	PE(21:0/0:0)	216.0632	198.0526	-	-	$C_{26}H_{55}NO_7P$	[M+H]+

Input Mass	Matched Mass	Delta	HG	C	DB	Abbreviation	M-sn1	M-sn1-H2O	M-sn2	M-sn2-H2O	Formula	Ion
675.6826	675.4960	.1866	PA	34	1	PA(12:0/22:1(13Z))	493.3289	475.3183	355.1881	337.1775	$C_{37}H_{72}O_8P$	[M+H]+
675.6826	675.4960	.1866	PA	34	1	PA(14:0/20:1(11Z))	465.2976	447.2870	383.2194	365.2088	$C_{37}H_{72}O_8P$	[M+H]+
675.6826	675.4960	.1866	PA	34	1	PA(14:1(9Z)/20:0)	467.3133	449.3027	381.2037	363.1931	$C_{37}H_{72}O_8P$	[M+H]+
675.6826	675.4960	.1866	PA	34	1	PA(16:0/18:1(9Z))	437.2663	419.2557	411.2507	393.2401	$C_{37}H_{72}O_8P$	[M+H]+
675.6826	675.4960	.1866	PA	34	1	PA(16:1(9Z)/18:0)	439.2820	421.2714	409.2350	391.2244	$C_{37}H_{72}O_8P$	[M+H]+

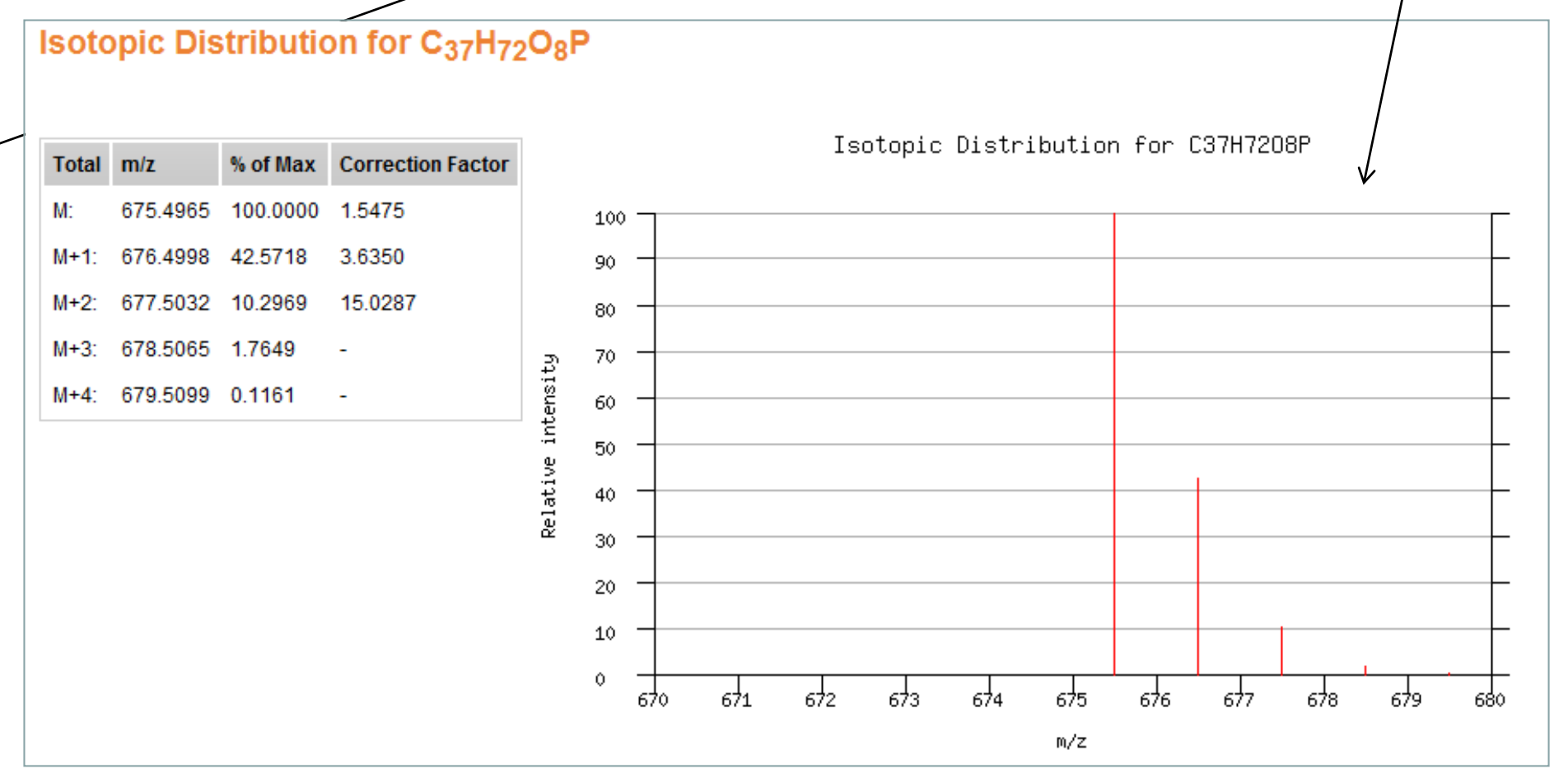
"On demand" generation of structure, systematic name and isotopic distribution profile



Glycerophospholipid Structure Generation

Download MDL MOL file containing only structure data or SDF file with structure and other associated data.

Abbreviation: PA(16:0/18:1(9Z))
Systematic Name: 1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphate
Formula (Neutral): $C_{37}H_{71}O_8P$
m/z: 675.496
Ion: [M+H]+
View Structure using MarvinView Applet JmolApplet ChemDraw (?)

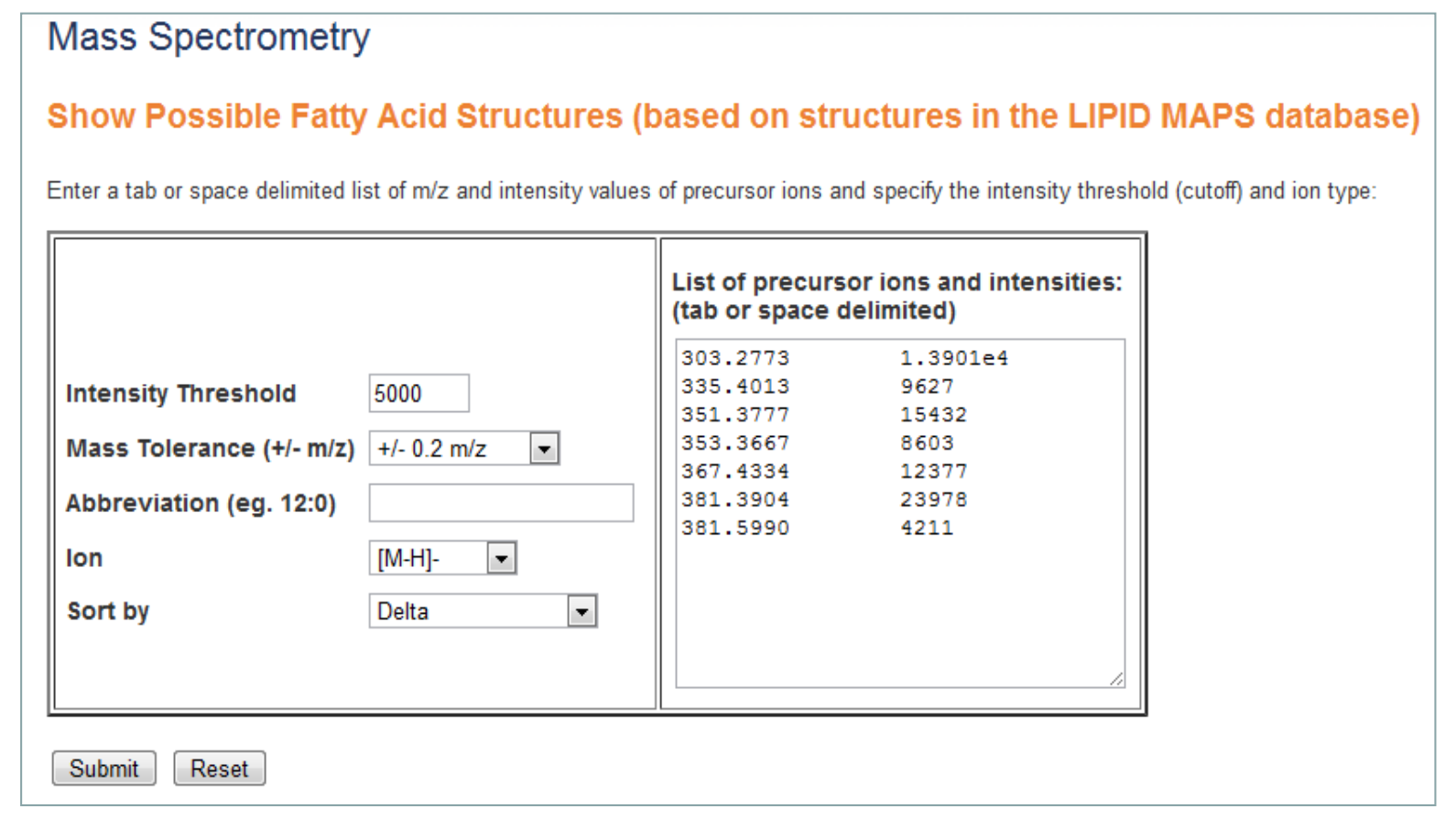


Isotopic Distribution for $C_{37}H_{72}O_8P$

Total	m/z	% of Max	Correction Factor
M:	675.4965	100.0000	1.5475
M+1:	676.4998	42.5718	3.6350
M+2:	677.5032	10.2969	15.0287
M+3:	678.5065	1.7649	-
M+4:	679.5099	0.1161	-

MS prediction using known lipids in the LIPID MAPS structure database (LMSD)

Search database of known Fatty acid structures (LMSD)



Mass Spectrometry

Show Possible Fatty Acid Structures (based on structures in the LIPID MAPS database)

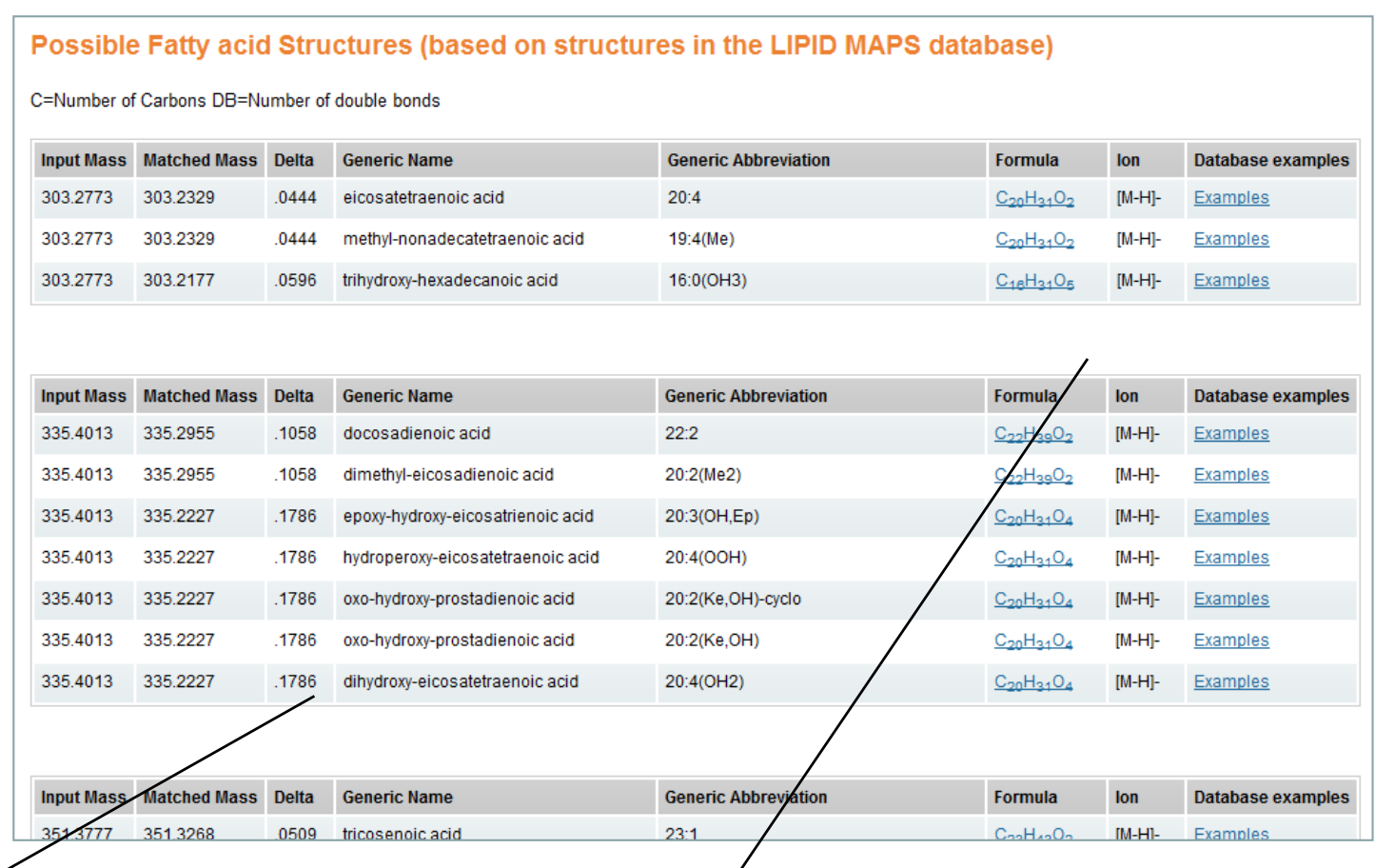
Enter a tab or space delimited list of m/z and intensity values of precursor ions and specify the intensity threshold (cutoff) and ion type:

Intensity Threshold: 5000
Mass Tolerance (+/- m/z): +/- 0.2 m/z
Abbreviation (eg. 12:0):
Ion: [M-H]-
Sort by: Delta

List of precursor ions and intensities: (tab or space delimited)

```
303.2773      1.3901e4
335.4013      9627
351.3777      15432
353.3667      8603
367.4334      12377
381.3904      23978
381.5990      4211
```

Results of MS search display "generic" matches

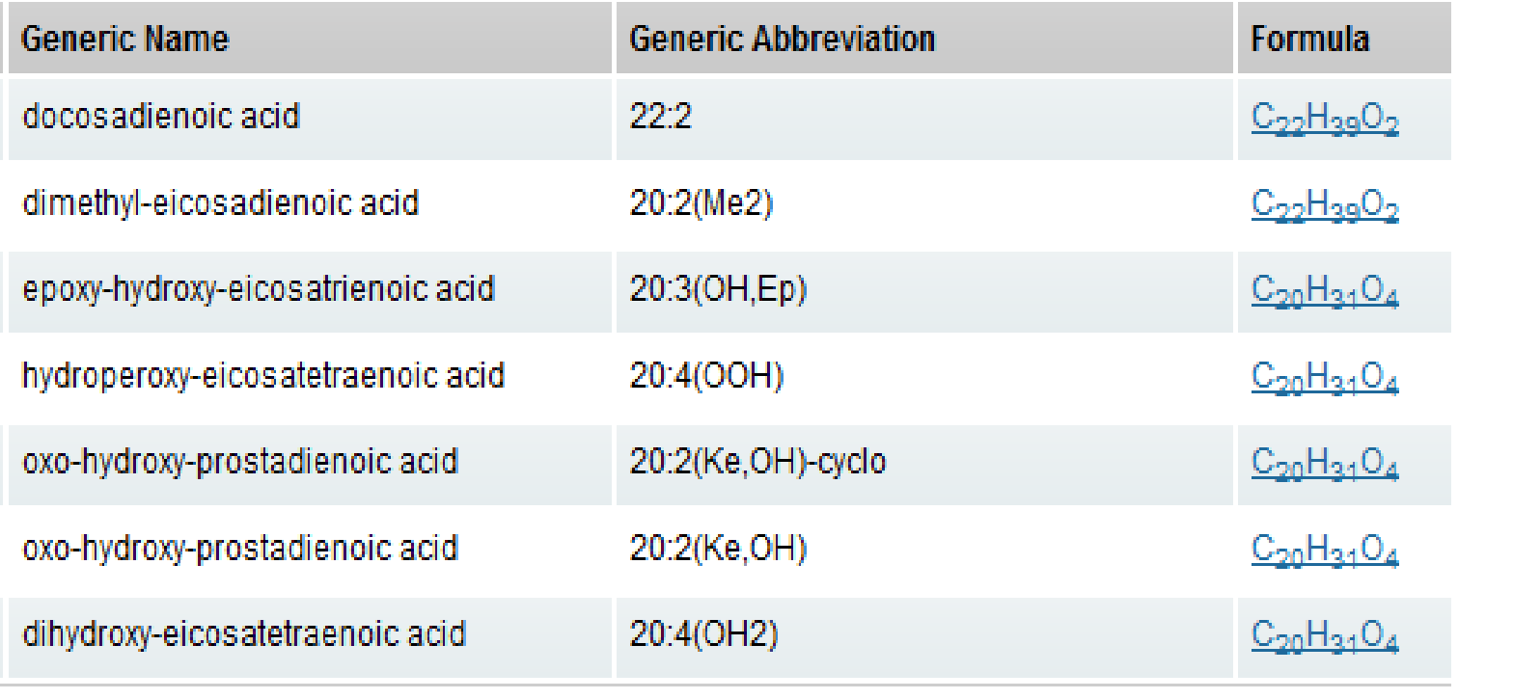


Possible Fatty acid Structures (based on structures in the LIPID MAPS database)

C=Number of Carbons DB=Number of double bonds

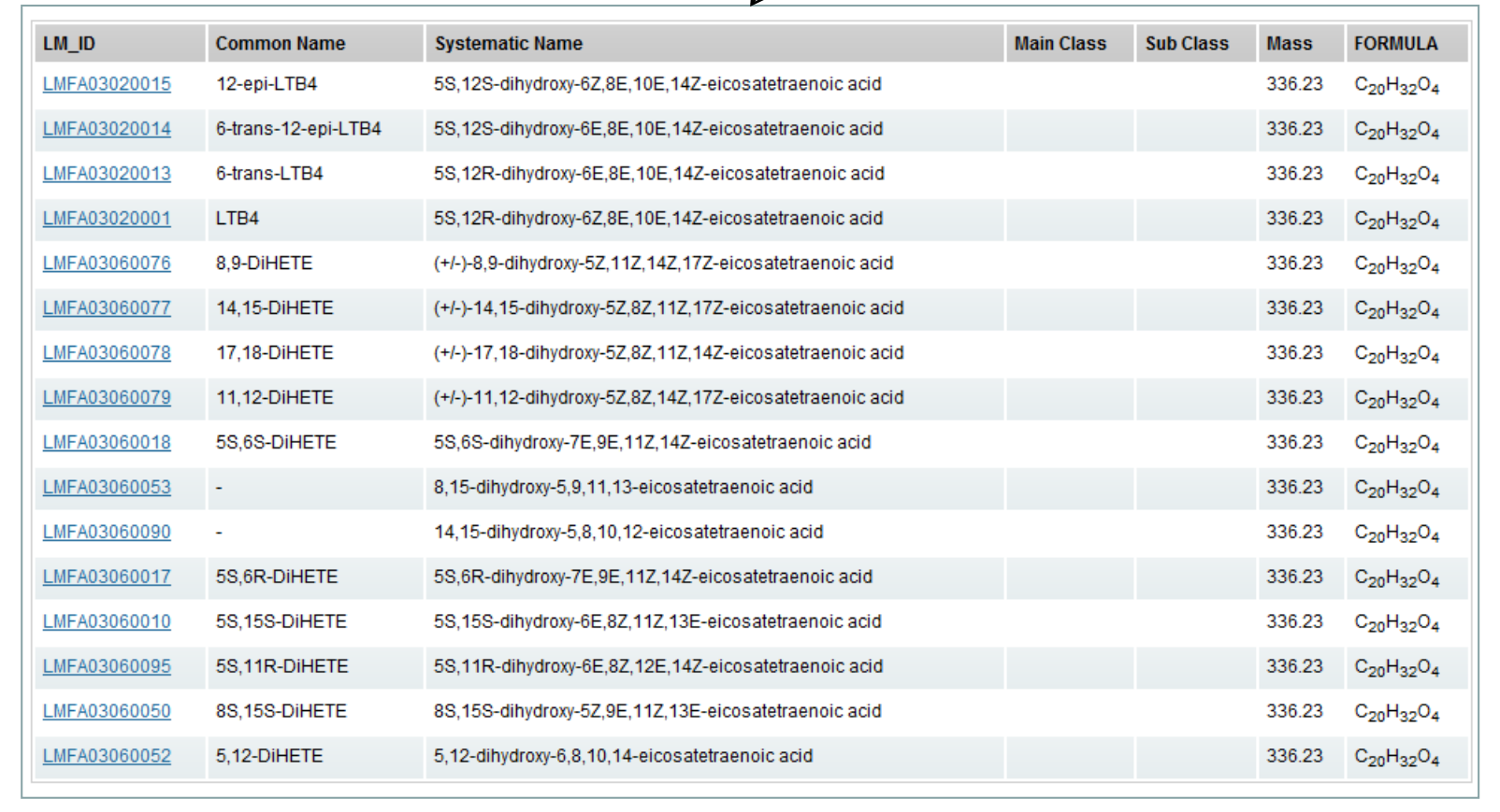
Input Mass	Matched Mass	Delta	Generic Name	Generic Abbreviation	Formula	Ion	Database examples
303.2773	303.2329	.0444	eicosatetraenoic acid	20:4	$C_{20}H_{31}O_2$	[M-H]-	Examples
303.2773	303.2329	.0444	methyl-nonadecatetraenoic acid	19:4(Me)	$C_{20}H_{31}O_2$	[M-H]-	Examples
303.2773	303.2177	.0596	trihydroxy-hexadecanoic acid	16:0(OH3)	$C_{16}H_{31}O_5$	[M-H]-	Examples

Input Mass	Matched Mass	Delta	Generic Name	Generic Abbreviation	Formula	Ion	Database examples
335.4013	335.2955	.1058	docosadienoic acid	22:2	$C_{22}H_{39}O_2$	[M-H]-	Examples
335.4013	335.2955	.1058	dimethyl-eicosadienoic acid	20:2(Me2)	$C_{22}H_{39}O_2$	[M-H]-	Examples
335.4013	335.2227	.1786	epoxy-hydroxy-eicosatrienoic acid	20:3(OH,Ep)	$C_{20}H_{31}O_4$	[M-H]-	Examples
335.4013	335.2227	.1786	hydroperoxy-eicosatetraenoic acid	20:4(OOH)	$C_{20}H_{31}O_4$	[M-H]-	Examples
335.4013	335.2227	.1786	oxo-hydroxy-prostadienoic acid	20:2(Ke,OH)-cyclo	$C_{20}H_{31}O_4$	[M-H]-	Examples
335.4013	335.2227	.1786	oxo-hydroxy-prostadienoic acid	20:2(Ke,OH)	$C_{20}H_{31}O_4$	[M-H]-	Examples
335.4013	335.2227	.1786	dihydroxy-eicosatetraenoic acid	20:4(OH2)	$C_{20}H_{31}O_4$	[M-H]-	Examples



Generic Name	Generic Abbreviation	Formula
docosadienoic acid	22:2	$C_{22}H_{39}O_2$
dimethyl-eicosadienoic acid	20:2(Me2)	$C_{22}H_{39}O_2$
epoxy-hydroxy-eicosatrienoic acid	20:3(OH,Ep)	$C_{20}H_{31}O_4$
hydroperoxy-eicosatetraenoic acid	20:4(OOH)	$C_{20}H_{31}O_4$
oxo-hydroxy-prostadienoic acid	20:2(Ke,OH)-cyclo	$C_{20}H_{31}O_4$
oxo-hydroxy-prostadienoic acid	20:2(Ke,OH)	$C_{20}H_{31}O_4$
dihydroxy-eicosatetraenoic acid	20:4(OH2)	$C_{20}H_{31}O_4$

"Generic" names and abbreviations



LM_ID	Common Name	Systematic Name	Main Class	Sub Class	Mass	FORMULA
LMFA03020015	12-epi-LTB4	5S,12S-dihydroxy-6Z,8E,10E,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03020014	6-trans-12-epi-LTB4	5S,12S-dihydroxy-6E,8E,10E,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03020013	6-trans-LTB4	5S,12R-dihydroxy-6E,8E,10E,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03020001	LTB4	5S,12R-dihydroxy-6Z,8E,10E,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060076	8,9-DiHETE	(+/-)-8,9-dihydroxy-5Z,11Z,14Z,17Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060077	14,15-DiHETE	(+/-)-14,15-dihydroxy-5Z,8Z,11Z,17Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060078	17,18-DiHETE	(+/-)-17,18-dihydroxy-5Z,8Z,11Z,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060079	11,12-DiHETE	(+/-)-11,12-dihydroxy-5Z,8Z,14Z,17Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060018	5S,6S-DiHETE	5S,6S-dihydroxy-7E,9E,11Z,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060053	-	8,15-dihydroxy-5,9,11,13-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060090	-	14,15-dihydroxy-5,8,10,12-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060017	5S,6R-DiHETE	5S,6R-dihydroxy-7E,9E,11Z,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060010	5S,15S-DiHETE	5S,15S-dihydroxy-6E,8Z,11Z,13E-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060095	5S,11R-DiHETE	5S,11R-dihydroxy-6E,8Z,12E,14Z-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060050	8S,15S-DiHETE	8S,15S-dihydroxy-5Z,9E,11Z,13E-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$
LMFA03060052	5,12-DiHETE	5,12-dihydroxy-6,8,10,14-eicosatetraenoic acid			336.23	$C_{20}H_{32}O_4$

Actual matches to mass and generic name in LMSD

LIPID MAPS Bioinformatics Core[a], UCSD, 9500 Gilman Dr, La Jolla, CA 92093; Department of Bioengineering[b], UCSD, 9500 Gilman Dr, La Jolla, CA, 92093