# Metabolite *In Silico* Identification Software (MetISIS)

Version 1.0.0

This computer software was prepared by Battelle Memorial Institute, hereinafter the Contractor, under Contract No. DE-AC05-76RL0 1830 with the Department of Energy (DOE). All rights in the computer software are reserved by DOE on behalf of the United States Government and the Contractor as provided in the Contract. You are authorized to use this computer software for non-commercial purposes (including activities performed for or on behalf of the U.S. Government) but you are not authorized to release or distribute it to others.

This material was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the United States Department of Energy, nor the Contractor, nor any or their employees, nor any jurisdiction or organization that has cooperated in the development of these materials, ***makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness or any information, apparatus, product, software, or process disclosed, or represents that its use would not infringe privately owned rights.***

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
*operated by*
BATTELLE
*for the*
UNITED STATES DEPARTMENT OF ENERGY
*under Contract DE-AC05-76RL01830*

# Metabolite *In Silico* Identification Software (MetISIS)

MetISIS is a software application that takes user observed MS/MS spectra and generates ranked lists of candidate identifications from a database of metabolites. MetISIS uses a database of molecular structures for which we have generated *in silico* spectra.

While MetISIS is still very limited in its capability, we share it in its current form in hopes it will be found useful. We are continuing to update the capability of MetISIS, and as we have new versions ready, we plan on adding these to the share site.

## Installing MetISIS

MetISIS is compiled as an MS Windows 32-bit application and runs on both 32-bit and 64-bit MS Windows operating systems. Unzipping the download creates a folder with the executable file MetISIS_32. This folder has two folders: Spectrum and SpectrumDB. The Spectrum folder has one user input file ObsSpectra.csv with one or more spectra. SpectrumDB has the spectra MetISIS needs—these should not be changed. Running MetISIS will create an output folder called ScreenResult.

## Running MetISIS

The MetISIS application is simple with Start, End, and About buttons. All filenames are hardcoded. MetISIS will read an example file ObsSpectra.csv from the Spectrum folder, and write an output file ScreenHits.csv to the ScreenResult folder. The download has one example file ObsSpectra.csv with 49 lipid molecules. The user who wants to try their spectra should backup the ObsSpectra.csv file, and make a new ObsSpectra.csv. The ObsSpectra.csv file format has (as the example shows) multiple spectra, each having the following content with each spectrum sequential:

- 1$^{st}$ line: one string (the example file has “[M+H]+” but the string is not used).
- 2$^{nd}$ line: one string that is written to the output result file. The output file has the ordinal numbers of spectra read from this file but the user may label or name the molecule on this line. If the result file is opened with Excel or other software that delimits data fields with commas, then it is better to not have commas in the string on this line.
- 3$^{rd}$ line: one real number, the observed precursor adduct m/z. Masses (charge +1 expected) should be < 1100 Da.
- 4$^{th}$ –n$^{th}$ line: pairs of real numbers, the spectra. First number is the m/z and the second the intensity (no normalization is required). Maximum number of spectral points per spectrum is currently set at 25k.
- n+1$^{th}$ line: the last line for each spectrum has four dollar signs ($$$$).

It takes a few seconds for MetISIS to run the 49 lipids in the example file, and then the output file ScreenHits.csv is available in the ScreenResult folder. Running the example input file makes one file to

view for format; the output has a header that hopefully is descriptive. The file is organized in the order the observed spectra were read (Test No). All the candidate identifications for one observed spectrum are grouped in consecutive rows. These rows are not ordered on rank. To make rank lists, use software such as Excel and sort on Test No as primary key and Score as secondary key (Score from largest to smallest). Score is the measure of similarity between the observed spectrum and the candidate spectrum. It is a simple Pearson R-square—where a higher score is a better match.

Observe that MetISIS searches for candidates within ± 500 ppm. The result file has a column PPM Off that lets the user remove candidates outside a lower mass margin.

## MetISIS Limitations

**Mass spectrometer.** MetISIS learned to make *in silico* spectra from lipid standards analyzed with a linear ion trap mass spectrometer (LTQ and LTQ Orbitrap, Thermo Scientific, San Jose, CA) and collision-induced dissociation. Similar instruments and fragmentation mechanisms may produce spectra that MetISIS can analyze. (It is in our future plan to add other instrument types.)

**LipidMaps.** MetISIS searches the LIPID Metabolites And Pathways Strategy (LIPID MAPS) Structure Database (LMSD) dated March 24, 2010 and available from www.lipidmaps.org. The database contains molecular data for >22,000 lipids. We have further reduced the database to lipids with masses < 1100 Da and to those lipids containing atoms from only {C, H, O, N, S, P}.

**Metabolite classes.** MetISIS needs to be trained with example data from more metabolite classes to make it useful for metabolites in general. The current shared version is trained with the following 13 classes/subclasses of lipid metabolites.

| Lipid Sub Class | Specie | Mass | Lipid Sub Class | Specie | Mass |
|---|---|---|---|---|---|
| Ceramide | d18:1/12:0 | 481.45 | Phosphatidylcholine | 14:0/14:0 | 677.50 |
| Ceramide | d18:1/17:0 | 551.53 | Phosphatidylcholine | 14:0/16:0 | 705.53 |
| Ceramide | d18:1/18:0 | 565.54 | Phosphatidylcholine | 16:0/16:0 | 733.56 |
| Ceramide | d18:1/20:0 | 593.57 | Phosphatidylcholine | 16:1/16:1 | 729.53 |
| Ceramide | d18:1/22:0 | 621.61 | Phosphatidylcholine | 17:0/17:0 | 761.59 |
| Ceramide | d18:1/24:0 | 649.64 | Phosphatidylcholine | 18:0/18:0 | 789.63 |
| Ceramide 1-phosphate | d18:1/12:0 | 561.42 | Phosphatidylcholine | 18:1/18:1 | 785.59 |
| Ceramide 1-phosphate | d18:1/16:0 | 617.48 | Phosphatidylcholine | 18:2/18:2 | 781.56 |
| Ceramide 1-phosphate | d18:1/18:1 | 643.49 | Phosphatidylcholine | 18:3/18:3 | 777.53 |
| Ceramide 1-phosphate | d18:1/24:0 | 729.60 | Phosphatidylcholine | 20:1/20:1 | 841.66 |
| Ceramide 1-phosphate | d18:1/8:0 | 505.35 | Phosphatidylcholine | 20:4/20:4 | 829.56 |
| Galactosyl(ß) ceramide | d18:1/12:0 | 643.50 | Phosphatidylcholine | 23:0/23:0 | 929.78 |
| Galactosyl(ß) ceramide | d18:1/16:0 | 699.56 | Phosphatidylethanolamine | 12:0/12:0 | 579.39 |
| Galactosyl(ß) ceramide | d18:1/24:1 | 809.67 | Phosphatidylethanolamine | 15:0/15:0 | 663.48 |
| Galactosyl(ß) ceramide | d18:1/8:0 | 587.44 | Phosphatidylethanolamine | 16:0/18:1 | 717.53 |
| Lactosyl(ß) ceramide | d18:1/12:0 | 805.56 | Phosphatidylethanolamine | 16:1/16:1 | 687.48 |
| Lactosyl(ß) ceramide | d18:1/16:0 | 861.62 | Phosphatidylethanolamine | 17:0/17:0 | 720.02 |
| Lactosyl(ß) ceramide | d18:1/24:0 | 973.74 | Phosphatidylethanolamine | 18:0/18:0 | 747.58 |
| Lactosyl(ß) ceramide | d18:1/8:0 | 749.49 | Phosphatidylethanolamine | 18:0/18:1 | 745.56 |
| Lysophosphatidylcholine | 14:0/0:0 | 467.30 | Phosphatidylserine | 12:0/12:0 | 623.38 |
| Lysophosphatidylcholine | 15:0/0:0 | 481.32 | Phosphatidylserine | 14:0/14:0 | 679.44 |
| Lysophosphatidylcholine | 16:0/0:0 | 495.33 | Phosphatidylserine | 16:0/18:2 | 759.47 |
| Lysophosphatidylcholine | 16:0/0:0 | 495.33 | Phosphatidylserine | 17:0/17:0 | 764.02 |
| Lysophosphatidylcholine | 17:0/0:0 | 509.35 | Phosphatidylserine | 18:0/18:0 | 792.07 |
| Lysophosphatidylcholine | 17:1/0:0 | 507.64 | Phosphatidylserine | 18:0/18:1 | 789.55 |
| Lysophosphatidylcholine | 18:0/0:0 | 523.36 | Phosphatidylserine | 18:0/18:2 | 787.54 |
| Lysophosphatidylcholine | 18:1/0:0 | 521.35 | Sphinganine | 17:0 | 287.28 |
| Lysophosphatidylethanolamine | 14:0/0:0 | 425.25 | Sphinganine 1-phosphate | 17:0 | 367.25 |
| Lysophosphatidylethanolamine | 16:0/0:0 | 453.29 | Sphingomyelin | d18:1/12:0 | 646.51 |
| Lysophosphatidylethanolamine | 18:0/0:0 | 481.32 | Sphingomyelin | d18:1/16:0 | 702.57 |
| Lysophosphatidylethanolamine | 18:1/0:0 | 479.30 | Sphingomyelin | d18:1/17:0 | 716.58 |
| Lysophosphatidylserine | 16:0/0:0 | 497.28 | Sphingomyelin | d18:1/18:1 | 728.58 |
| Lysophosphatidylserine | 18:0/0:0 | 525.31 | Sphingomyelin | d18:1/24:1 | 812.68 |
| Lysophosphatidylserine | 18:1/0:0 | 523.60 | | | |

MetISIS is currently in training with more lipid classes and the future plan is to extend MetISIS to metabolites in general.

**Metabolite Adducts and Charges.** The shared MetISIS version is only trained to identify [M+H]+ adducts. MetISIS will also match observed spectra to candidate [M+NH4]+ adducts. These candidates should be assumed false positives as explained below.

## Expected Rank Results

The table below shows scores from training MetISIS with lipid standards. Our experience in preliminary biological studies is that MetISIS assigns scores, similar to those in the table, to good quality spectra (containing only one species).

| Lipid | Score | Lipid | Score |
|---|---|---|---|
| Phosphatidylcholine | 0.995 | Ceramide | 0.998 |
| Phosphatidylcholine | 0.999 | Ceramide | 0.980 |
| Phosphatidylcholine | 1.000 | Ceramide | 0.975 |
| Phosphatidylcholine | 1.000 | Ceramide | 0.943 |
| Phosphatidylcholine | 1.000 | Ceramide | 0.942 |
| Phosphatidylcholine | 0.999 | Ceramide | 0.964 |
| Phosphatidylcholine | 0.997 | Ceramide 1-phosphate | 0.967 |
| Phosphatidylcholine | 1.000 | Ceramide 1-phosphate | 0.941 |
| Phosphatidylcholine | 0.999 | Ceramide 1-phosphate | 0.957 |
| Phosphatidylcholine | 0.998 | Ceramide 1-phosphate | 0.958 |
| Phosphatidylcholine | 0.991 | Ceramide 1-phosphate | 0.972 |
| Phosphatidylcholine | 0.997 | Galactosyl(ß) ceramide | 0.925 |
| Phosphatidylcholine (e) | 0.626 | Galactosyl(ß) ceramide | 0.977 |
| Phosphatidylcholine (e) | 0.610 | Galactosyl(ß) ceramide | 0.968 |
| Phosphatidylcholine (p) | 0.988 | Galactosyl(ß) ceramide | 0.945 |
| Phosphatidylcholine (p) | 0.990 | Lactosyl(ß) ceramide | 0.856 |
| Phosphatidylethanolamine | 0.999 | Lactosyl(ß) ceramide | 0.861 |
| Phosphatidylethanolamine | 0.994 | Lactosyl(ß) ceramide | 0.907 |
| Phosphatidylethanolamine | 1.000 | Lactosyl(ß) ceramide | 0.876 |
| Phosphatidylethanolamine | 0.999 | Lysophosphatidylcholine | 0.964 |
| Phosphatidylethanolamine | 0.993 | Lysophosphatidylcholine | 0.916 |
| Phosphatidylethanolamine | 0.999 | Lysophosphatidylcholine | 0.992 |
| Phosphatidylethanolamine | 0.999 | Lysophosphatidylcholine | 0.951 |
| Phosphatidylserine | 0.993 | Lysophosphatidylcholine | 0.959 |
| Phosphatidylserine | 0.948 | Lysophosphatidylcholine | 0.807 |
| Phosphatidylserine | 0.981 | Lysophosphatidylcholine | 0.849 |
| Phosphatidylserine | 0.994 | Lysophosphatidylcholine | 0.843 |
| Phosphatidylserine | 0.998 | Lysophosphatidylethanolamine | 0.940 |
| Phosphatidylserine | 0.985 | Lysophosphatidylethanolamine | 0.771 |
| Phosphatidylserine | 0.989 | Lysophosphatidylethanolamine | 0.532 |
| Sphinganine | 0.975 | Lysophosphatidylethanolamine | 0.778 |
| Sphinganine | 0.915 | Lysophosphatidylethanolamine | 0.889 |
| Sphinganine 1-phosphate | 0.913 | Lysophosphatidylserine | 0.853 |
| Sphingomyelin | 0.944 | Lysophosphatidylserine | 0.873 |
| Sphingomyelin | 0.960 | Lysophosphatidylserine | 0.905 |
| Sphingomyelin | 0.976 | | |
| Sphingomyelin | 0.911 | | |
| Sphingomyelin | 0.910 | | |

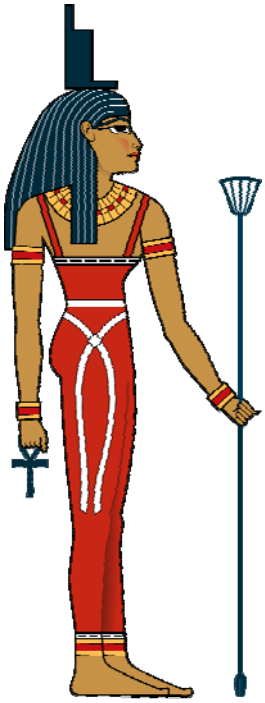
Most candidates from lipid classes that are currently not in the MetISIS training set, score low, but some may score higher than the positive candidate. The table below shows the correct candidate PS(16:0/0.0) falling below a few Fatty Acyls and Sterol Lipids. There are two reasons why a rank list like this is

suspicious: first, lipids from two classes above the correct candidate have not yet been trained in MetISIS, and second, the difference in the scores between the correct candidate and the top rank is only 0.006596—not enough to separate candidates, assuming no other information is available.

| Test No | User Lbl | Obs Mass | Score | DB No | Adduct Mass | PPM Off | LM ID | Class | Specie | Mass | Formula |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | 16-0_LPS_ | 498.33 | 0.884901 | 41 | 498.5250054 | 391.1648 | LMFA01010032 | Fatty Acyls [FA] | Lacceroic acid | 480.4906313 | C32H64O2 |
| 31 | 16-0_LPS_ | 498.33 | 0.884901 | 17401 | 498.3253193 | 9.3928 | LMST03020303 | Sterol Lipids [ST] | (6R)-1alpha,25-di | 480.2909452 | C27H44O5S |
| 31 | 16-0_LPS_ | 498.33 | 0.884901 | 17402 | 498.3253193 | 9.3928 | LMST03020304 | Sterol Lipids [ST] | (6S)-1alpha,25-di | 480.2909452 | C27H44O5S |
| 31 | 16-0_LPS_ | 498.33 | 0.884795 | 3176 | 498.5250054 | 391.1648 | LMFA07010001 | Fatty Acyls [FA] | Cetyl palmitate | 480.4906313 | C32H64O2 |
| 31 | 16-0_LPS_ | 498.33 | 0.88203 | 17403 | 498.3253193 | 9.3928 | LMST03020305 | Sterol Lipids [ST] | (6RS,24R)-24,25-d | 480.2909452 | C27H44O5S |
| 31 | 16-0_LPS_ | 498.33 | 0.881128 | 17600 | 498.3947195 | 129.856 | LMST03020512 | Sterol Lipids [ST] | (22E,24E,24bE)-1a | 480.3603454 | C32H48O3 |
| 31 | 16-0_LPS_ | 498.33 | 0.878305 | 8198 | 498.2833085 | 93.70472 | LMGP03050002 | Glycerophospholipids | PS(16:0/0:0) | 497.2753685 | C22H44NO9P |

We have performed a test of MetISIS and compared it to another algorithm (results are forthcoming). Overall, MetISIS appears to have high sensitivity and specificity, but we also recognize that MetISIS is currently only applicable to phospholipids observed in positive mode mass spectrometry—an obvious shortcoming if other metabolite classes are of interest.

## Contact Information

We would appreciate feedback on the MetISIS performance and would appreciate hearing of any problems. Due to a time constraint to distribute MetISIS, we have only added minimal error checking in the code at this time.

Email: Lars.Kangas@pnnl.gov