

Lipid Data Analyzer for Sphingolipids

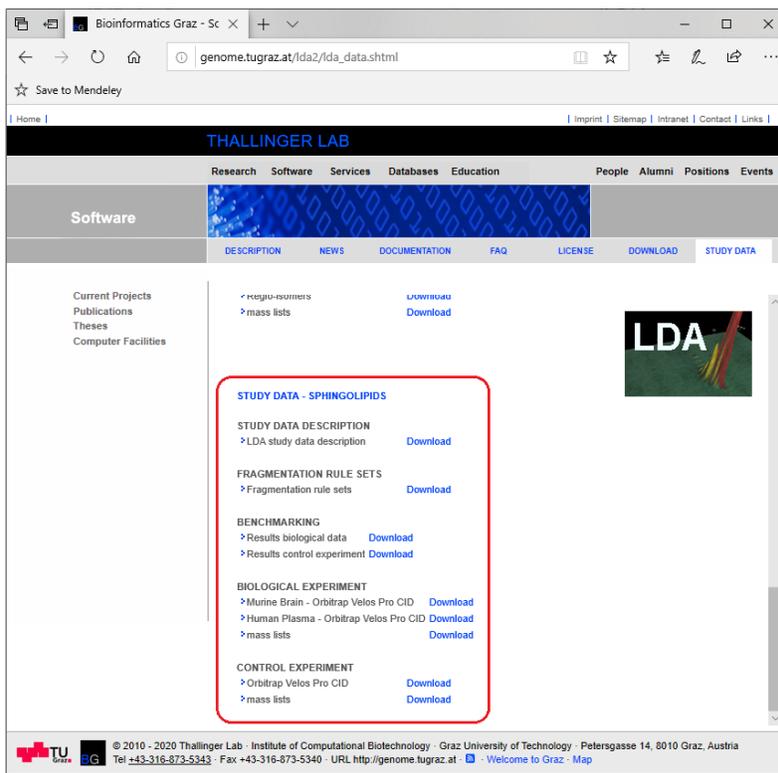
Download and Data Instructions

1	Download instructions.....	1
1.1	Instructions for study data download	1
2	Data description of benchmarking	2
3	Data visualization.....	3
3.1	Visualization by LDA	3
4	Available data	6
4.1	Biological data	7
4.2	Control experiment	7

1 Download instructions

Study data are available at:

http://genome.tugraz.at/lda2/lda_data.shtml



1.1 Instructions for study data download

Study data is categorized in 'BENCHMARKING', 'BIOLOGICAL EXPERIMENT', and 'CONTROL EXPERIMENT'. Furthermore, the fragmentation rules and this description is provided in the sections 'FRAGMENTATION RULE SETS', and 'STUDY DATA DESCRIPTION', respectively.

The screenshot shows the Thallinger Lab website. The main navigation bar includes 'Home', 'Imprint', 'Sitemap', 'Intranet', 'Contact', and 'Links'. Below this is a secondary navigation bar with 'Research', 'Software', 'Services', 'Databases', 'Education', 'People', 'Alumni', 'Positions', and 'Events'. The 'Software' section is highlighted, and a sub-menu shows 'DESCRIPTION', 'NEWS', 'DOCUMENTATION', 'FAQ', 'LICENSE', 'DOWNLOAD', and 'STUDY DATA'. The 'DOWNLOAD' link is selected, leading to a page with a sidebar on the left containing 'Current Projects', 'Publications', 'Theses', and 'Computer Facilities'. The main content area lists several categories with download links: 'regio-summers' and 'mass lists' (both with 'Download' links); 'STUDY DATA - SPHINGOLIPIDS' with a sub-link 'STUDY DATA DESCRIPTION' and a 'Download' link; 'FRAGMENTATION RULE SETS' with a sub-link 'Fragmentation rule sets' and a 'Download' link; 'BENCHMARKING' with sub-links 'Results biological data' and 'Results control experiment', each with a 'Download' link; 'BIOLOGICAL EXPERIMENT' with sub-links 'Murine Brain - Orbitrap Velos Pro CID', 'Human Plasma - Orbitrap Velos Pro CID', and 'mass lists', each with a 'Download' link; and 'CONTROL EXPERIMENT' with sub-links 'Orbitrap Velos Pro CID' and 'mass lists', each with a 'Download' link. A red box highlights the 'Download' link for 'Murine Brain - Orbitrap Velos Pro CID', with a red note next to it: 'Download raw data and results obtained on the murine brain sample'. The footer contains copyright information for 2010-2020 Thallinger Lab, contact details for the Institute of Computational Biotechnology at Graz University of Technology, and a 'Welcome to Graz' message.

The data is further stratified in subcategories, such as e.g. ‘Murine Brain – Orbitrap Velos Pro CID’ for ‘BIOLOGICAL EXPERIMENT’. Next to these subcategories, links for downloading data sets are provided. The categories ‘BIOLOGICAL EXPERIMENT’ and ‘CONTROL EXPERIMENT’ provide separate downloads for the mass lists used as input for LDA.

Data from various experiments contain typically three folders: massLists (quantitation Excels used as input for LDA), and positive and negative for the polarities. For the control experiment, the positive and negative folders contain folders for Mix1 and Mix2. In most of the cases, the folders positive and negative (and Mix1 and Mix2) contain MS runs, comments, and evaluation results. Further details about the individual files are given in chapters 2 and 4, and their visualization in chapter 3.

2 Data description of benchmarking

The final evaluation results are stored in the following Excel files:

Brain_LDA_MS-Dial_positive.xlsx and Brain_LDA_MS-Dial_negative.xlsx (biological)

LDA-MS-DIAL_positive_standards.xlsx and LDA-MS-DIAL_negative_standards.xlsx (control experiment).

These Excel files consist of a ‘Summary’ tab containing the final results, and five or ten more tabs for the evaluation results of the individual MS runs. Detected lipid species and lipid molecular species are counted in two ways: adduct sensitive, where every species/adduct combination is one hit; adduct insensitive, where every detected species is one hit. E.g. Cer d18:0/n18:0 is found in protonated and sodiated form; in the first evaluation scheme, these are counted as two hits, in the second one, this corresponds to one hit only. Tables in the manuscript are based on the adduct insensitive numbers. For the biological experiment, data is evaluated for sensitivity (percent of correctly identified species relative to the total number of manually identified species) and positive predictive value (percent of correctly identified species relative to the total number of predictions made by the corresponding software) on the lipid species and lipid molecular species level. For control experiment 1, the sensitivity is evaluated in three aspects: 1) lipid species detection rate; 2) detection rate of correct chain combinations; 3) detection rate of correct chain combinations without reporting additional wrong chain combinations.

There are five or ten more Excel files provided, where the names end by “comp.xlsx”. These files contain details about the individual comparisons of LDA and MS-DIAL. The “Comment” columns contain detailed reasons for counting a hit as false positive, or why LDA missed a hit.

Furthermore, there is an ‘export’ folder that contains the original MS-DIAL output.

In the folders where the individual comparisons of LDA and MS-DIAL are stored, raw data in the MS-DIAL input format (abf) are provided. The rest of the files in these folders are files generated by MS-DIAL.

The original Thermo RAW files and the LDA results can be found in the corresponding ‘BIOLOGICAL EXPERIMENT’ and ‘CONTROL EXPERIMENT’ sections.

3 Data visualization

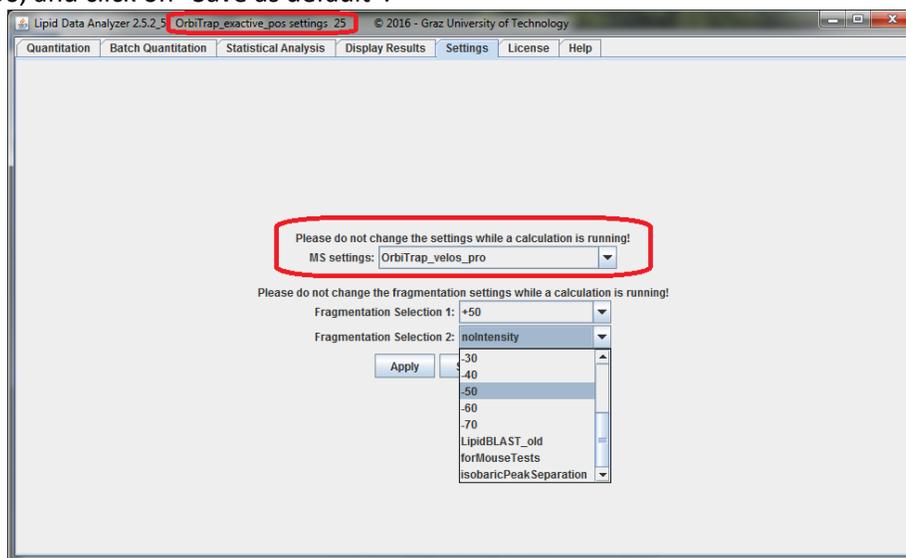
This chapter covers data description and visualization of all experiments, except for the benchmarking study which is described in chapter 2.

A complete data set (one polarity, one collision energy, one mix) consists of five MS runs and some txt or Excel files containing comments and/or evaluation results. One MS run comprises typically of three files, i.e., data in its vendor format (raw), a ‘chrom’ directory (data in LDA format), and an Excel file containing the results produced by LDA.

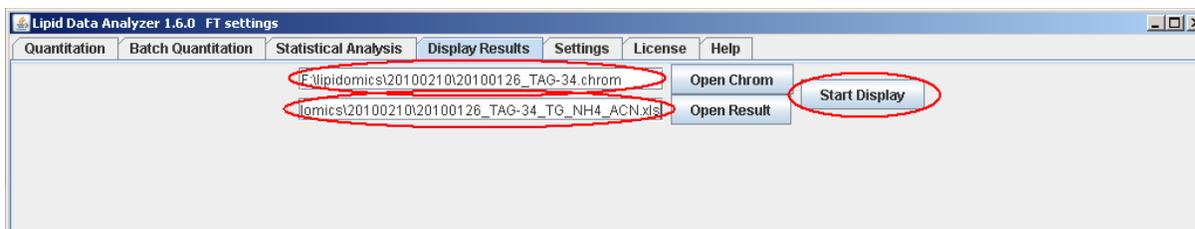
3.1 Visualization by LDA

In order to install the LDA, use the provided installation wizards. During the installation process, it does not matter which MS platform you choose, since this can be changed easily in the application itself. Usually the default settings in the wizard should be sufficient.

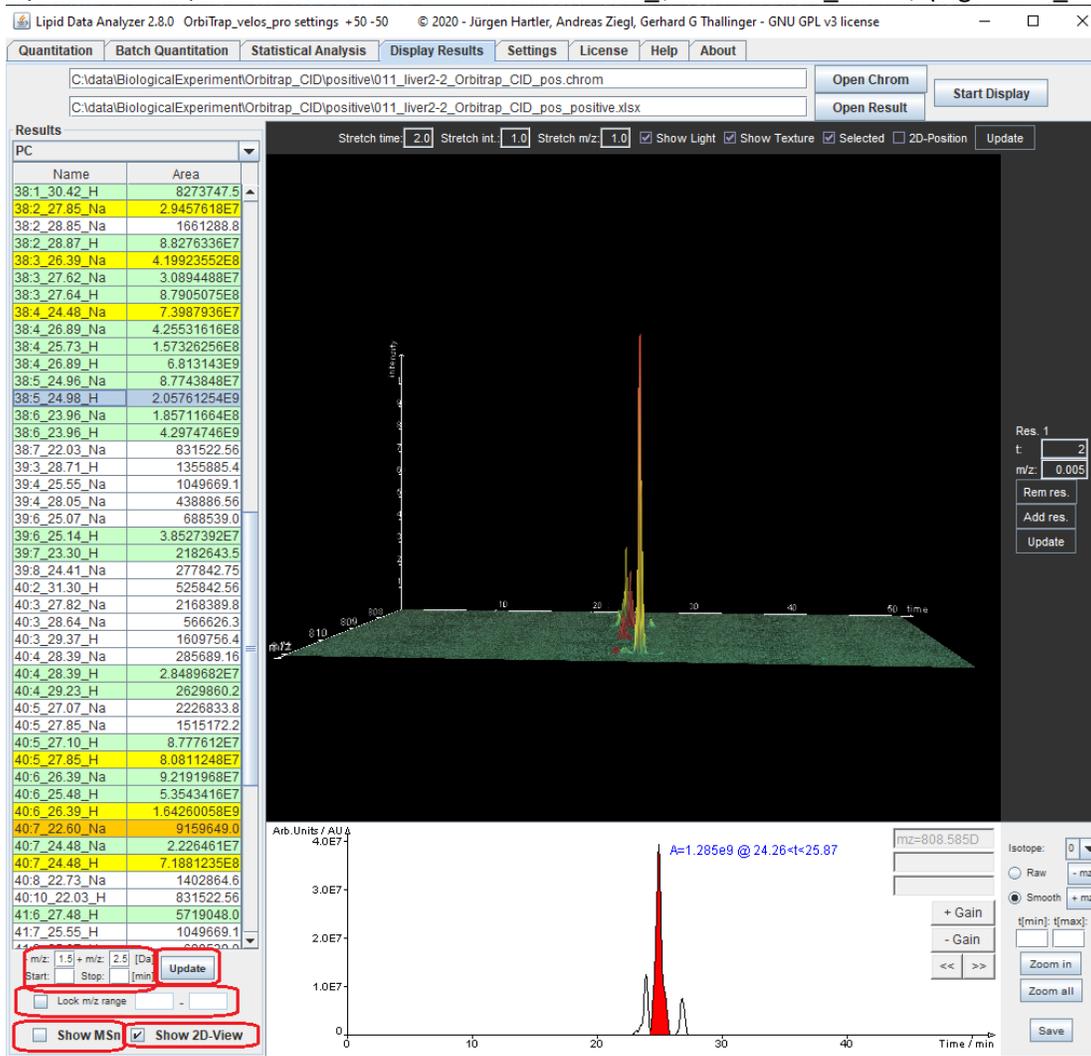
In order to visualize data of a single MS run, start the LDA, click on “Settings” tab, select Orbitrap_velos_pro, select +50 and -50, and click on “Save as default”:



Then, click on the “Display Results” tab, select the corresponding chrom and LDA Excel result file, and click on “Start Display”:



A menu on the left side will appear now containing a select box at the top, where you can switch between the different lipid subclasses. Then a table follows, whereupon the name of the analyte is in the first column, and the integrated mass-chromatographic area in the second. The display name consists primarily of “\$name\$: \$double_bonds\$ \$retention_time\$” (e.g. d42:1_29.76 lipid species name). If there are more than one adduct/modification, the modification name is added with _\$modification_name\$ (e.g. d42:1_29.76_H).



The color encoding of the table entries has the following meaning:

- White: MS¹ identification without any MSⁿ evidence (not relevant for this publication).
- Green: MS¹ identification verified by MSⁿ spectra.
- Yellow: The MSⁿ algorithm detected more than one lipid subclass in the MS¹ peak → the peak was split on the MS¹ level in a part for the one and a part for the other lipid subclass. The split is based on information derived from MSⁿ spectra.
- Orange: The MSⁿ algorithm detected more than one lipid subclass in this peak, but a split on the MS¹ level is not possible. The intensities of the intact MS¹ peak are split according to the distinct fragments

of MSⁿ spectra. This way of splitting is rather unreliable, since usually fragment intensities do not correlate with MS¹ abundances.

- Blue: Current selection for display.

The “Show MSⁿ” checkbox switches from the lipid species to the lipid molecular species view of the table where possible:

Name	Area
20:1_18:0_30.42_H	8273747.5
38:2_27.85_Na	2.9457618E7
38:2_28.85_Na	1661288.8
18:1/20:1_28.87_H	8472244.0
18:0/20:2_28.87_H	7.9804088E7
18:0/20:3_26.39_Na	4.19923552E8
18:0/20:3_27.62_Na	3.0894488E7
21:1_17:2_27.64_H	9.5518848E7
18:0/20:3_27.64_H	7.835319E8
38:4_24.48_Na	7.3987936E7
18:0/20:4_26.89_Na	4.25531616E8
22:4_16:0_25.73_H	1.3572973E7
20:4_18:0_25.73_H	3.7298596E7
18:1/20:3_25.73_H	1.06454688E8
18:0/20:4_26.89_H	6.813143E9
18:1/20:4_24.96_Na	8.7743848E7
16:0/22:5_24.98_H	5.31407136E8
20:0/18:5_24.98_H	4.6825152E8
18:1/20:4_24.98_H	1.05795392E9
16:0/22:6_23.96_Na	1.85711664E8
16:0/22:6_23.96_H	4.2974746E9
38:7_22.03_Na	831522.56
39:3_28.71_H	1355885.4

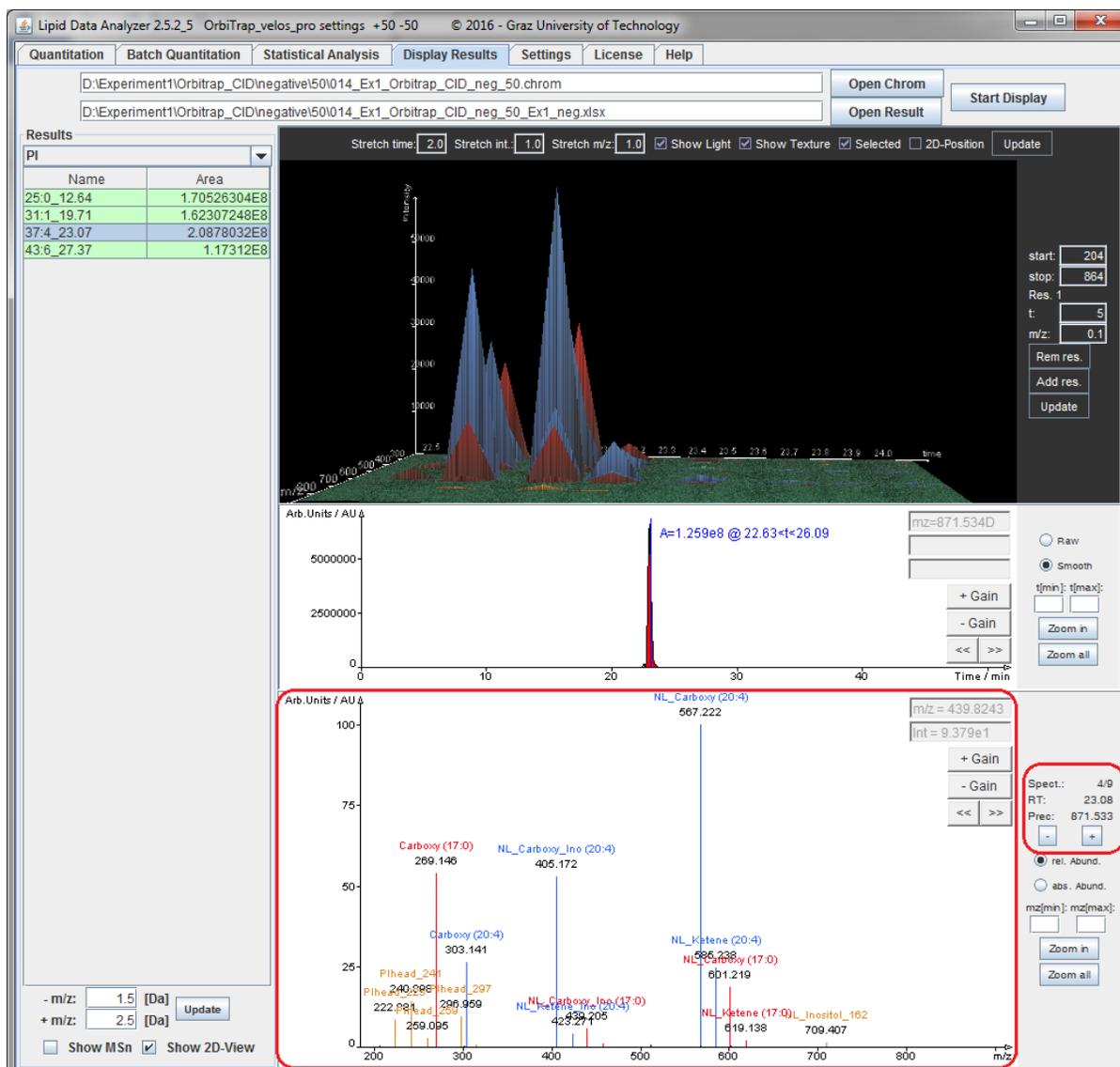
The table displays now information about the obtained structure according to Liebisch, G. et al. Shorthand notation for lipid structures derived from mass spectrometry. J Lipid Res 54, 1523-1530 (2013). I.e., the sum of carbon atoms and double bonds in the chains if no structural information can be obtained (e.g. 38:5), the fatty acid chains separated by an underscore if fatty acids are detectable (e.g. 20:4_18:0), or the fatty acid chains separated by a slash if the position of the fatty acid chains is assignable (e.g. 18:1/20:4). Attention: For the manual inspection of MSⁿ, there is a difference in the displayed annotation of the spectra, if the lipid species consists of more than one lipid molecular species:

- The lipid species view is selected: The fragments of all detected lipid molecular species are annotated.
- The lipid molecular species view is selected: Only the fragments of the particularly selected lipid molecular species are annotated.

When a row of the table is activated by the left mouse button, the peaks are displayed in the 3D and the chromatogram viewer. When a row is activated by the right mouse button, a popup appears:

38:3_27.64_H	8.7905075E8
38:4_24.48_Na	7.3987936E7
38:4_26.89_Na	4.25531616E8
38:4_25.73_H	1.57326256E8
38:4_26.89_H	6.813143E9
38:5_24.96_Na	8.7743848E7
38:5_24.98_H	2.06761264E9
38:6_23.96_Na	
38:6_23.96_H	
38:7_22.03_Na	
39:3_28.71_H	
39:4_25.55_Na	
39:4_28.05_Na	
39:6_25.07_Na	
39:6_25.14_H	
39:7_23.30_H	
39:8_24.41_Na	277842.75
40:2_31.30_H	525842.56

By clicking “Show MS/MS”, MS/MS spectra are shown as annotated by the used MSⁿ decision rules at the time of quantitation (see figure on next page).



The big red box shows an annotated spectrum, and the small red box shows the buttons +/- for scrolling through spectra. Further details regarding LDA visualization options can be found in the user manual which is accessible in the "Help" tab of the LDA.

4 Available data

This chapter covers all datasets except for the benchmarking study which is explained in chapter 2. Data sets or assays consist of five runs, and 1 or more txt or Excel files containing results of detailed analysis, comments about information content that can be derived from the various adduct ions. Each MS run consists of a 'chrom' directory (raw data in LDA format), an Excel file containing the results produced by LDA, and the data in its vendor format. The following chapters explain in detail the amount of additional information provided for each study.

4.1 Biological data

Each data set in biological data consists of five MS runs, and several Excel files containing manual evaluation results for both positive and negative ion mode. These files contain at least one “\$TISSUE\$_SpeciesDetectable ...” Excel file, reporting the identified species, e.g. Brain_SpeciesDetectable_MSMS_Orbitrap_CID.xlsx, and one file containing “FP” in the name, reporting the identified false positives on the MS2 level (there is no evaluation of the MS¹ level, since this study covers fragmentation spectra only). Furthermore, an Excel document starting with “FNs ...” is available, reporting missed identifications, and several Excel documents containing the lipid subclass and the “species” in the file name, reporting details which lipid species and lipid molecular species were found in the individual MS runs.

4.2 Control experiment

Each data set consists of 5 MS runs for each ion mode and mix and several text files containing details about the detected species, missed species and false positives. Additional comments are available providing general information about the individual subclasses/adducts. These files are located in the directories named positive and negative.