

Software for Mass Spectrometry-based Lipidomics

Subjects: Biochemistry & Molecular Biology

Contributor: Nils Hoffmann, Gerhard Mayer, Canan Has, Dominik Kopczynski, Fadi Al Machot, Dominik Schwudke, Robert Ahrends, Katrin Marcus, Martin Eisenacher, Michael Turewicz

Mass spectrometry (MS) is a state-of-the-art analytical technology, which enables the rapid and consistent identification and quantification of lipids in lipidomics, metabolites in metabolomics and proteins in proteomics for biomedical and biochemical research purposes.

Keywords: lipidomics ; bioinformatics ; data format ; database ; mass spectrometry

1. Introduction

Mass spectrometry (MS) is a state-of-the-art analytical technology, which enables the rapid and consistent identification and quantification of lipids in lipidomics, metabolites in metabolomics and proteins in proteomics for biomedical and biochemical research purposes ^[1]. Through the technological advances achieved during the past twenty years, main performance parameters were improved, such as mass accuracy and sensitivity. MS has become the analytical method of choice for many omics disciplines. All MS-based omics technologies share the following general workflow: (i) sample separation, (ii) analysis by a separation technology such as liquid chromatography (LC), hydrophilic interaction liquid chromatography (HILIC), reversed phase liquid chromatography (RPLC), supercritical fluid chromatography (SFC), gas chromatography (GC) or capillary electrophoresis (CE), (iii) mass spectrometric measurements supported by different ionization principles, e.g., via electrospray (ESI), electron ionization (EI), desorption electrospray ionization (DESI) for 'matrix-assisted laser desorption and ionization' (MALDI), (iv) separation and detection of the ions by the m/z values in the mass analyzer applying several physical principles and (v) storage of MS spectra, where the signal intensities are proportional to the abundance of the molecular species. However, applied omics workflows are comprised of several specific customizations to be well suited for the investigated biomolecule class and the associated analytical question.

Lately, ion mobility spectrometry (IMS) has gained a lot of attraction as a method of separating ions in the gas phase ^[2]. In IMS, ions are brought into interaction with an inert collision gas using static or modulated electric field gradient configurations to achieve ion separation and selection. An ion's retention behavior in the IMS separator is determined by its average rotational collisional cross section (CCS), such that more compact ions tend to migrate faster toward the outlet of the IMS separator by exhibiting fewer collisions. Further, its behavior is influenced by the interaction of the ion with the superimposed electric field and effective waveform, which can either filter (FAIMS) ions with specific mobility, separate ions in an electric field gradient within a drift tube (DTIMS) or separate ions into ion packets by a traveling wave electric field within stacked ring ion guide (TWIMS).

For the further characterization of a given molecule in a targeted lipidomics workflow for the validation and quantification of lipids, specific precursor m/z values and select potential fragment m/z values (transitions in an inclusion list) are tracked using robust and comparably inexpensive triple-quad MS instruments in selective reaction monitoring (SRM) mode, which allows for the identification and quantitation of lipids on the class level. Orbitrap-type or time-of-flight (TOF) MS instruments with a higher mass resolution and the ability to perform a full-scan acquisition in parallel reaction monitoring (PRM) mode for selected precursors, measuring all fragment ions simultaneously, can be used for targeted lipidomics to achieve a deeper MS fragment coverage, allowing for species or subspecies identification.

In untargeted lipidomics workflows for discovery applications, no previous inclusion list is provided, thus requiring MS instruments that can operate in a data-dependent acquisition (DDA) or data-independent acquisition (DIA) mode to obtain a full-scan precursor and fragment mass spectra of either top- k m/z signals with the highest intensity or all ions contained in predefined m/z windows. Such experiments are often performed on instruments with high mass resolutions to further reduce ambiguities caused by isobaric lipids.

Tandem mass spectrometric experiments (MS^2) are applied to gain further insights into the lipid structure and various fragmentation methods are applicable to record precursor-specific fragment spectra. However, collision-induced dissociation (CID) is the most widely established approach. Identification software is applied to identify molecules by comparison of generated MS^2 spectra with theoretical fragment spectra or with reference spectra from a database. The quantification of molecules is usually performed using the corresponding precursor mass spectra but may also be performed on selected MS^2 fragments. Higher-level fragmentation series for identification and quantification are also applicable, where the mass spectrometer selects MS^2 fragment ions for further fragmentation (MS^n). Finally, the resulting data, i.e., raw MS^1 and MS^n spectra and chromatographic retention time (RT), drift time or collisional cross section, scan

polarity, collision energies and corresponding metadata such as MS device settings, are stored in vendor-specific data formats.

2. Software for Lipid Identification from Mass Spectrometry

The recent development of lipid identification tools has aimed to propel the rapidly emerging field of lipidomics by improving the quality and performance of applied algorithms, while integrating novel separation techniques and high-resolution mass spectrometers. Researchers reviewed a total of 31 openly available software tools for lipidomics data processing and identification that were published between 2006 and the end of 2021. Researchers evaluate the usability of common data formats and, specifically, of PSI standard data formats as either input or output formats and their support for at least one of the lipidomics workflows (see **Table 1** for reference).

Table 1. Overview of software for lipid identification from mass spectrometry. Abbreviations: U: Untargeted, T: Targeted, C: Chromatography, CE: Capillary Electrophoresis, IM: Ion Mobility, DI: Direct Infusion (Shotgun), I: Imaging. \$: targeted includes Selected Reaction and Multiple Reaction Monitoring (MRM), untargeted includes DDA and DIA approaches. *: Only the most important ones relevant to this research. All tools use some form of configuration file format, e.g., text-based (TXT) or other formats for libraries or fragmentation rules. Workflow assignment designates the primary workflow a tool was designed for and this was stated by the authors; others may be available. Researchers use direct infusion as a more generic synonym for what is usually referred to as 'shotgun lipidomics'. Comma-separated values (CSV) is a tabular, spreadsheet-like format. If tab characters are used as separators, the format is TSV. Hypertext markup language (HTML) is a format viewable with an internet browser. XLSX: MS office XML-based spreadsheet format. MSP: NIST mass spectral library format. MGF: Mascot Generic Format. BLIB: Binary mass spectral library format. PDF: Portable Document Format. #: rule-based validation often includes spectral scores, ratios and thresholds, scores denote spectral similarity functions, such as the commonly used dot product/cosine variants. Remarks: (1) The software is no longer available. (2) Lipid class separation chromatography, e.g., HILIC or supercritical fluid chromatography. (3) XCMS input recommended, LIPID MAPS class assignment of suspect ions. (3) Software is provided as a web application without further information. (4) Supports phospholipids only. (5) XCMS input recommended, LIPID MAPS class assignment of suspect ions. (6) After release 3.0, LipidMatch is available as LipidMatch Flow (latest version 3.5, but without source code). (7) Supports oxidized phospholipids only. (8) Identification and quantification use other tools' methods. (9) The source code is provided for download, but no code license is defined.

Workflow \$	Name	Handling	MS *	Identification #	Quant	Input	Output	Last Release	Open-Source	License
T	LIMSA	C, DI	MS ¹ , MS ²	Compound/Fragment library	yes	XLSX, CSV, HTML	NA	2006	NA (1)	GPL v3
T	LipidomeDB	DI, C	MS ¹ , MS ²	m/z Library + Transitions + rule-based	yes	XLSX	XLSX, HTML	2019	no	NA
T	LipidQuant	C (2)	MS ¹	m/z library + rule-based	yes	TXT	XLSX	2021	yes	CC-BY 4
U	ALEX and ALEX 123	DI	MS ¹ , MS ² , MS ³	Manual	no	manual input of parameters	HTML	2017	no	NA
U	Greazy (4)	C, DI	MS ¹ , MS ²	Fragment/Spectral Library + score	no	vendor, mzML	mzTab (via LipidLama)	2022	yes	Apache v2
U	LDA2	C	MS ¹ , MS ²	Rule-based	yes	mzML, TXT	XLSX, mzTab-M	2021	yes	GPL v3
U	LipidBlast	C	MS ¹ , MS ²	Spectral Library + score	no	MSP, MGF, XLSX	MGF, XLSX	2014	yes	CC-BY
U	LipidDex	C	MS ¹ , MS ²	Spectral Library + rule-based	yes	MGF, mzXML, CSV	CSV	2018	yes	MIT
U	LipidFinder	C	MS ¹	Rule-based, LMSD	no	CSV, JSON (5)	PDF	2021	yes	MIT
U	LipidHunter (4)	C, DI	MS ¹ , MS ²	Rule-based	yes	mzML, XLSX, TXT	XLSX, HTML, TXT	2020	yes	GPL v2, Proprietary
U	LipidIMMS	C, IM	MS ¹ + CCS, MS ²	CCS Library + Spectral Library + score	no	MSP, MGF	CSV, HTML	2020	no	NA
U	LipidMatch (6)	C, I, DI	MS ¹ , MS ² , MSE/DIA	Compound/Fragment library + rule-based	yes	CSV, MS2 (ProteoWizard)	CSV	2020	yes	CC BY 4.0
U	LipidMiner	C	MS ¹ , MS ²	Compound/Fragment library + rule-based	yes	raw	XLSX, CSV	2014	no	NA

Workflow \$	Name	Handling	MS *	Identification #	Quant	Input	Output	Last Release	Open-Source	License
U	LipidMS	C	MS ¹ , MS ² , MSE/DIA	Compound/Fragment library + rule-based	yes	mzXML, CSV	CSV	2022	yes	GPL v3
U	Lipid-Pro	C	MSE/DIA	Compound/Fragment library	yes	CSV	XLSX, TXT	2015	no	Proprietar
U	LipidXplorer	DI	MS ¹ , MS ² , MS ³	Rule based	no	mzML (MS ¹ + MS ²)	CSV, HTML	2019	yes	GPL v2
U	LiPydomics	C, IM	MS ¹	CCS Library + <i>m/z</i> Library + HILIC RT Library + rule-based	yes	CSV	XLSX	2021	yes	MIT
U	LIQUID	C	MS ¹ , MS ²	Spectral Library + rule-based	yes	RAW, mzML	TSV, mzTab, MSP	2021	yes	Apache v.
U	LOBSTAHS	C	MS ¹	Spectral Library + rule-based	yes	mzML, mzXML, mzData, CSV	XLSX, CSV	2021	yes	GPL v3
U	LPPTiger (7)	C	MS ¹ , MS ²	Spectral Library + score	yes	mzML, XLSX, TXT	XLSX, HTML	2021	yes	GPL v2, Proprietar
U	MassPix	I	MS ¹	<i>m/z</i> Library + rule-based	no	imzML	CSV	2017	yes	NA
U	MS-DIAL 4	C, CE, IM	MS ¹ , MS ² , MSE/DIA	Spectral Library + rule-based	yes	vendor, mzML	CSV, mzTab-M, XLSX	2022	yes	GPL v3
U	MZmine 2	C	MS ¹ , MS ²	Spectral Library + rule-based	yes	vendor, mzML, mzXML, mzData, CSV, mzTab, XML	CSV, mzTab, XML	2019	yes	GPL v2
U	XCMS	C	MS ¹ , MS ²	Spectral Library + score	yes	mzML, mzXML, netCDF	CSV	2021	yes	GPL v2
T + U	LipidCreator and Skyline	C	MS ¹ , MS ² , MSE/DIA	Fragment/Spectral Library + score (8)	yes (8)	vendor, mzML (MS ¹ + MS ²)	XLSX, CSV, BLIB	2021	yes	MIT
T + U	LipidPioneer	C	MS ¹ , MS ²	Compound/ <i>m/z</i> Library (8)	yes (8)	XLSX	XLSX	2017	yes (9)	NA
T + U	LipidQA	DI	MS ¹ , MS ²	Spectral Library + score	yes	vendor (Thermo, Waters)	CSV	2007	NA (1)	NA
T + U	LipoStar	C, IM	MS ¹ , MS ² , MSE/DIA	Compound/Fragment library + rule-based validation	yes	vendor	CSV	2022	no	Proprietar
T + U	LipoStarMSI	DI, I	MS ¹ , MS ²	Spectral Library + rule based	yes	vendor (Bruker, Waters), imzML	CSV	2020	no	Proprietar
T + U	SmartPeak	C	MS ¹ , MS ²	Transitions + rule-based	yes	mzML, CSV	mzTab, XML, CSV	2022	yes	MIT
T + U	Smfinder	C	MS ¹ , MS ²	Spectral Library + score	yes	mzML, mzXML	XLSX, TXT	2020	yes (9)	NA

Researchers categorized the tools by supported workflow (targeted, untargeted or both), sample handling (separation, e.g., chromatography, ion mobility, direct infusion, imaging), MS level, summarizing targeted, selected ions under MS¹, MS² for shotgun and DDA approaches and MSE/DIA for data-independent approaches, based on their own claims in their primary publications or documentation.

Concerning lipid identification, researchers broadly distinguish between tools that use either a rule-based or a library-based identification approach.

Rule-based tools must describe at least precursor ion *m/z*, MS² fragments and (relative) fragment intensity ranges for lipid class, species or subspecies identification. In order to reduce the chance for false-positive identifications, these approaches often also apply further validation rules, such as fragment signal intensity ratios that must fall within certain bounds. However, these rule-based approaches can be customized to also allow for identification on a more precise lipid

structure level if the necessary data is available. In principle, these approaches are very flexible and allow for the query of spectra for certain patterns that are indicative of specific lipid species. This makes them applicable to targeted, as well as untargeted, analysis.

Library-based approaches use either in-silico generated MS² spectra for lipids derived from their structural representation or experimentally acquired and post-processed spectra. To assign a putative identity to measured lipid mass spectra, a variant of the dot product score or other related vector scores is often used [3][4].

Researchers further indicate whether tools support quantitative output, such as intensities, areas, relative or absolute quantities or if they only support qualitative lipid identification output. For these tools to be included in larger processing workflows, the supported data formats for input and output are crucial. In the mass spectrometry and lipidomics field specifically, researchers can distinguish between text (human readable) and binary file formats. The latter are often the raw data vendor formats, but can also include local database files, such as the blib format for mass spectral libraries or the common sqlite database format. Within text-based formats, researchers can distinguish structured ones that follow a specific schema for MS data, such as the Mascot Generic Format (MGF), NIST Mass Spectrum format (MSP), MS2 [5] or mzTab(-M) and semi-structured ones, such as CSV, JSON or XLSX, where the latter is a compressed XML format. XML-based formats are well-adapted to be machine readable and validatable and are used in the PSI format mzML, as well as its predecessors, mzXML [6] and mzData [7]. TXT formats are generally only weakly structured but remain human-readable.

Maintenance, accessibility and reusability are important factors in being able to create and maintain reproducible processing pipelines from openly available tools. Researchers therefore also captured the date of the last release for each tool with a granularity of one year and whether it is available under an explicit open-source license, and if so, under which one specifically. This is also an important aspect for the original authors of a tool, as sustainable development and maintenance of bioinformatics software through a lack of continued funding is still an issue. Open access to the software can help in building up a community around it, where maintenance and further development can be shared between different stakeholders. Researchers did not specifically record whether a tool's source code is available via a source code repository platform such as GitHub or GitLab, but generally recommend that for open-source software, since these platforms will make the source code available for the foreseeable future.

Lastly, researchers list the programming languages that were used to develop the tool. This can have an impact on operating system platform independence and may make reuse of the software easier for certain user demographics, e.g., MS EXCEL and VBA macros may simplify usage by non-bioinformaticians but have clear limits to the Windows platform and limit integrability into non-UI driven workflows.

2.1. Targeted Workflow

LIMSA [8][9] supports data from both LC separation, as well as direct infusion workflows. In a first step, vendor data needs to be converted to the NetCDF format using the authors proprietary but free of charge tool, SECD, which is then used to export MS data to LIMSA via EXCEL. LIMSA itself is implemented in C++ as an EXCEL add-in and provides peak finding, identification, isotopic correction and absolute quantification based on calibration lines and labeled internal standards. Unfortunately, researchers were not able to find a publicly available version of the software.

LipidomeDB [10][11] is a web application for the processing of direct infusion and differential ion mobility MS lipidomics data. It requires a user login but is otherwise free to use. LipidomeDB supports isotopic correction and absolute quantification via class-specific labeled lipid standards and linear calibration curves. Input data needs to be provided in XLSX format and can be exported after identification and quantification as XLSX and HTML.

LipidQuant [12] is a tool for quantitative lipidomics in lipid class separation workflows, such as HILIC or SFC coupled to MS, based on EXCEL and Visual Basic for Applications (VBA). It supports input of *m/z* and sample-wise quantity data in TXT or generally tabular formats from vendor software. It includes an extensible built-in database of lipid species, organized by lipid class, and performs type II isotopic correction and absolute quantification using class-specific, heavy labeled (deuterated) internal lipid standards. Output is available from the XLSX worksheet.

2.2. Untargeted Workflow

ALEX 123 [13] is an online database that provides comprehensive fragmentation information on 430,000 lipid molecules from 47 lipid classes across five different lipid categories. Output of ALEX 123 is provided in HTML format. In combination with LDA2, it was used for lipid and lipid fragment identification in LC-MS/MS data. Alternatively, ALEX [14] can be used for lipid identification on a species level from high-resolution FTMS data. The source codes of ALEX and ALEX 123 are not publicly available.

Greazy [15] is well-integrated with the ProteoWizard tool suite and supports both chromatography-MS as well as DI data. It generates a search space of phospholipids and theoretical MS² spectra based on user input. Experimental MS² spectra are searched against the phospholipids in the search space with adjustable precursor mass tolerance. The match score is computed based on a combination of hypergeometric distribution and intensity score, considering the number of observed

fragments for each lipid. The lipid spectrum matches are filtered based on density estimation and the hits above the score threshold are reported in mzTab 1.0 format.

Lipid Data Analyzer 2 (LDA2) ^{[14][15]} supports untargeted LC-MS/MS lipidomics workflows and is implemented in JAVA. It accepts the following input formats for MS data: raw, .d, wiff, chrom and mzXML. It requires additional quantitation files (XLSX) with lipid class/species to mass/adduct mass association and additional expected RTs for each experiment. In LDA2, custom platform and ionization energy-specific fragmentation rule sets for lipid class and scan species level fragment identification can be defined. Identification and quantification results are stored in XLSX, CSV, mzTab 1.0 and most recently, mzTab-M 2.0.

LipidBlast ^{[16][17][18]} is a suite of XLSX/Visual Basic for Applications (VBA) macros that can generate in-silico tandem MS libraries for lipid identification with other tools, such as NIST's MS Search application. Input formats are MSP, MGF and XLSX, while output can be generated in MGF and XLSX formats. It is not actively developed any longer, but its libraries have been integrated into MS-DIAL.

LipidDex ^[19] is also implemented in JAVA. It uses in-silico fragmentation templates and lipid-optimized MS² spectral matching to identify and track lipid species in LC-MS/MS experiments. It can calculate peak purity and determine co-isolation and co-elution of isobaric lipids and is able to remove ionization artifacts. It reads data in MGF or mzXML formats and saves identification results in CSV tables.

LipidFinder ^{[20][21][22]} is a Python tool and web application available from the LIPID MAPS website that supports untargeted identification of lipids in LC-MS data, using XCMS for initial feature finding and custom filter and post-processing steps specifically tailored to lipidomics. Input formats are those that are also supported by XCMS, but specifically CSV and JSON, to transfer feature data and configuration settings to the application. LipidFinder supports the generation of reports in PDF, XLSX and CSV formats.

LipidHunter ^[23] identifications are based on (glycero-)phospholipidomics MS² spectra measured by RPLC-MS/MS or direct infusion methods, integrating with LIPID MAPS for bulk lipid search. It supports mzML as an input format from LC-MS/MS and data-dependent shotgun acquisitions. Input files need to be split into an MS¹-only file, covering survey scans for faster processing, and a complete file that contains MS¹ and MS² scans. LipidHunter extracts fragment ions based on a user-definable configuration and links MS² fragment information to parent ions that are identified against the LIPID MAPS database. It finally performs a lipid species assignment based on their product ions and additional rules. LipidHunter reports quantification and identification results in HTML, CSV and XLSX.

LipidIMMS Analyzer ^{[24][25]} is a web application for lipid identification in chromatography ion mobility workflows. It uses an internal database of MS¹, CCS, RT and MS² information and applies a weighted composite scoring to assign the final identification. It accepts data in MSP and MGF formats and supports output in CSV and HTML.

LipidMatch ^[26] supports LC-MS, imaging and direct infusion workflows based on an extensive in-silico MS² fragmentation library including 56 different lipid types. It uses a rule-based approach for lipid identification against the precursor and fragment *m/z* values, including definable adducts, and it is implemented in R. DDA as well as DIA data are supported through peak picking with tools such as MZmine or XCMS. LipidMatch accepts input in CSV (feature tables) or MS² (MS/MS data) format and provides annotated and identified results down to the subspecies fatty acyl level. It exports identification results in CSV format. LipidMatch Flow converts vendor file formats with msConvert on the fly.

LipidMiner ^[27] supports LC-MS/MS DDA data and uses the LIPID MAPS structure database as its library for lipid identification using a rule-based approach. It is implemented in Python and C# and provides input from Thermo raw files. Output is provided in XLSX and CSV formats.

LipidMS ^[13] is an R package that supports the processing of high-resolution, DIA-MS data. Due to the missing direct relation between the precursor and fragments in DIA, the package applies a score to assess the co-elution of both for grouping, based on fragment and ion intensity rules that allow annotation on species, molecular subspecies (fatty acyl) and structural species (FA position) level. Input may be provided in mzXML or CSV. Output is available as R objects, which can be easily converted and exported into CSV and other tabular formats.

Lipid-Pro ^[28] is another tool that supports DIA LC-MS/MS data. Implemented in C#, it uses a lipid compound and fragment library and applies matching rules to identify precursor fragment associations based on retention time-aligned, pre-processed data. Input can be provided in CSV format, while output is available as XLSX or TXT.

LipidXplorer ^{[29][30]} supports DI-MS lipidomics workflows regardless of the lipid category, implemented in Python. It transfers filtered and averaged representative spectra (from all scans based on the measurement settings of the data) into a master scan. The master scan is then searched against the fragmentation rules per class and per mode as provided by query scripts written in Molecular Fragmentation Query Language (MFQL), which is inspired by the SQL database query language. The tool currently supports Thermo raw and mzML files as well as text file-based import (CSV for MS¹ and DTA for MS², in v1.2.7) as input files and generates comma-separated output files. The output file can be programmed by

MFQL and usually reports lipid species found with mass, chemical formula, identification error, lipid name, isobaric species, if any, along with precursor and fragment ion intensities per sample (CSV).

LiPydomics ^[31] is a Python tool for HILIC ion mobility MS lipidomics data analysis. It uses a custom experimental database with *m/z* and CCS values for 45 lipid classes and HILIC retention times for 23 lipid classes. CCS prediction and HILIC retention time prediction for lipids that are not contained in the experimental database are realized by applying machine learning to the experimental database reference values. Identification is performed using a rule-based approach on *m/z*, RT and CCS values. LiPydomics accepts CSV files as input and provides results in XLSX format.

LIQUID ^[32] supports identification of lipids from LC-MS/MS experiments with a customizable library and adaptable scoring model that includes quartiles of fragment intensities. The library covers over 30,000 lipid targets in nine distinct lipid categories, 29 lipid classes and 85 subclasses, sourced from LIPID MAPS and extended with additional lipids. It is implemented in C# and supports input in Thermo Fisher raw format and mzML. Processing results can be exported in CSV, mzTab or MSP formats.

LOBSTAHS ^[33] is implemented in R for the identification of lipids, oxidized lipids and oxylipin biomarkers in LC-MS data. It uses XCMS and the R/Bioconductor package CAMERA ^[34] for feature detection and aggregation and validates potential lipid features against an internal *m/z* library of lipid species adducts using a rule-based approach based on adduct order of intensity. Input is therefore supported in all formats that XCMS supports. Output can be exported in XLSX or CSV formats.

For oxidized phospholipids, LPPTiger ^[22] is an option for data-dependent LC-MS/MS data. It is implemented in Python and uses in-silico generated spectral libraries together with a composite score based on individual similarity, rank, fingerprint, isotope matching and specificity scores. It reads data in mzML, XLSX and TXT formats as input (MSP for the library format) and outputs as XLSX and HTML.

MassPix ^[35] is an R library for the analysis of imaging-MS lipidomics data. It uses an MS¹ *m/z* library for rule-based identification. It reads imzML format as input and annotates deisotoped *m/z* values against its internal generated library. Identified results can be exported in CSV format.

MS-DIAL 4 ^{[36][37]}, written in JAVA, supports chromatography, CE and ion mobility workflows. It applies a spectral library search approach, based on a MS fragment library of 177 lipid subclasses. MS-DIAL 4 performs peak picking, alignment annotation and quantification. Identification combines scoring and a rule-based approach that is guided by a decision tree and provides different levels of confidence. As input formats, multiple vendor formats and mzML are supported, while outputs can be written in CSV, XLSX and mzTab-M. MS-DIAL also supports retention time prediction and offers comprehensive visualizations.

MZmine 2 ^[38] is a modular software for untargeted, chromatography-based metabolomics, with support for lipid species identification using spectral libraries and rules for annotation. It is implemented in JAVA and offers to read input from a variety of vendor formats as well as from open formats as input and it is also able to export identification and intensity data in common spreadsheet and tabular formats and supports mzTab for reading and writing. The upcoming MZmine 3 will also support mzTab-M.

XCMS ^{[39][40]} is a generic R/Bioconductor library for mass spectrometry feature finding and grouping and has no dedicated support for lipid identification. It uses a spectral library-based approach for feature identification, but other packages may provide other functionality more tailored for lipids. XCMS supports LC-MS/MS data in mzML, mzXML and netCDF formats and outputs feature tables in CSV, XLSX or other formats supported by the R ecosystem.

2.3. Targeted and Untargeted Workflow

The final batch of tools support the analysis of targeted and semi-targeted or untargeted lipidomics data.

LipidCreator ^{[41][42]}, together with Skyline ^[43], is primarily designed for targeted lipidomics analysis, but through Skyline's support for DIA analysis, can also be applied for untargeted workflows. LipidCreator is used to create transition lists and spectral libraries for more than 60 lipid classes, either using predefined libraries for common species and tissues or by manual selection of lipid classes, head groups and fatty acyl parameters. Transitions and a spectral library derived from the in-silico transition list can be transferred to Skyline to be used with its peak/transition detection and integration and its spectral matching features. All major vendor formats are supported, as well as mzML for input. Results can be exported in XLSX and CSV formats, while spectral libraries are exported in the open BLIB format.

LipidPioneer ^[44] is an EXCEL template implemented in VBA supporting more than 60 lipid classes, including oxidized ones. It allows the generation of custom lipid inclusion lists based on sum formulas of adduct masses for use in targeted and untargeted workflows. These can then be used by other software for lipid identification, such as MZmine, MS-DIAL or Greazy, or for Quality Assurance (QA) and Quality Control (QC) applications. LipidPioneer supports export in any format supported by EXCEL, e.g., CSV or EXCEL.

LipidQA ^[45] supports both targeted and untargeted workflows for DI-MS. It is implemented in Visual C++ and uses a fragment ion and lipid chemical formula database to perform spectral matching for identification. Absolute quantitation with calibration curves is also supported. LipidQA can read data in Thermo and Waters vendor formats and provides its results in CSV format.

LipoStar ^[46], implemented in C#, supports data from chromatographic separation and ion mobility for DDA and DIA workflows. It uses a compound and fragment library and rule-based validation for the identification of lipids. LipoStar reads vendor MS data and supports the exporting of results in the CSV format.

LipoStarMSI ^[47] is LipoStar's sibling software for direct infusion and imaging MS lipidomics. It uses a spectral library and rule-based approach for lipid identification. LipoStarMSI is also implemented in C# and can read vendor formats of Bruker and Waters as well as the open imzML format. Output is exported in CSV format.

SmartPeak ^[48] uses OpenMS ^[49] at its core and supports absolute quantitation in targeted and semi-targeted workflows. It is implemented mainly in C++ and implements MRM-specific peak integration and feature selection on top of established OpenMS methods. SmartPeak's primary input format is mzML, while transitions, parameters and sample sequence information are provided in CSV format. Results can be exported in mzTab, XML and CSV formats.

Smfinder ^[50] has parts that are implemented in Python and some parts that are implemented in R. It supports targeted, untargeted and ¹³C labeling workflows. Lipid identification is performed based on plausible sum formulas first, with subsequent validation using a spectral library. The untargeted workflow uses XCMS for feature detection. Smfinder supports mzML and mzXML as input data formats. Results can be exported in XLSX and TXT formats.

Out of the 31 tools for lipid identification researchers reviewed, 6 of 31 (>19%) did not provide a release version that could help to ensure reproducibility when authors want to compare their software to those of others. Eight of 31 tools (>25%) had no explicit license defined. Just as many, but not necessarily the same ones, did not provide the source code in an openly accessible way.

References

1. Züllig, T.; Trötzmüller, M.; Köfeler, H.C. Lipidomics from Sample Preparation to Data Analysis: A Primer. *Anal. Bioanal. Chem.* 2020, 412, 2191–2209.
2. Paglia, G.; Smith, A.J.; Astarita, G. Ion Mobility Mass Spectrometry in the Omics Era: Challenges and Opportunities for Metabolomics and Lipidomics. *Mass Spectrom. Rev.* 2021.
3. Stein, S.E.; Scott, D.R. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.* 1994, 5, 859–866.
4. Yilmaz, Ş.; Vandermarliere, E.; Martens, L. Methods to Calculate Spectrum Similarity. In *Proteome Bioinformatics*; Keerthikumar, S., Mathivanan, S., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2017; pp. 75–100. ISBN 978-1-4939-6740-7.
5. McDonald, W.H.; Tabb, D.L.; Sadygov, R.G.; MacCoss, M.J.; Venable, J.; Graumann, J.; Johnson, J.R.; Cociorva, D.; Yates, J.R., III. MS1, MS2, and SQT—Three Unified, Compact, and Easily Parsed File Formats for the Storage of Shotgun Proteomic Spectra and Identifications. *Rapid Commun. Mass Spectrom.* 2004, 18, 2162–2168.
6. Oliver, S.G.; Paton, N.W.; Taylor, C.F. A Common Open Representation of Mass Spectrometry Data and Its Application to Proteomics Research. *Nat. Biotechnol.* 2004, 22, 1459–1466.
7. Orchard, S.; Montechi-Palazzi, L.; Deutsch, E.W.; Binz, P.-A.; Jones, A.R.; Paton, N.; Pizarro, A.; Creasy, D.M.; Wojcik, J.; Hermjakob, H. Five Years of Progress in the Standardization of Proteomics Data 4th Annual Spring Workshop of the HUPO-Proteomics Standards Initiative April 23–25, 2007 Ecole Nationale Supérieure (ENS), Lyon, France. *Proteomics* 2007, 7, 3436–3440.
8. Haimi, P.; Uphoff, A.; Hermansson, M.; Somerharju, P. Software Tools for Analysis of Mass Spectrometric Lipidome Data. *Anal. Chem.* 2006, 78, 8324–8331.
9. Haimi, P.; Chaithanya, K.; Kainu, V.; Hermansson, M.; Somerharju, P. Instrument-Independent Software Tools for the Analysis of MS-MS and LC-MS Lipidomics Data. *Methods Mol. Biol.* 2009, 580, 285–294.
10. Zhou, Z.; Marepally, S.R.; Nune, D.S.; Pallakollu, P.; Ragan, G.; Roth, M.R.; Wang, L.; Lushington, G.H.; Visvanathan, M.; Welti, R. LipidomeDB Data Calculation Environment: Online Processing of Direct-Infusion Mass Spectral Data for Lipid Profiles. *Lipids* 2011, 46, 879–884.
11. Fruehan, C.; Johnson, D.; Welti, R. LipidomeDB Data Calculation Environment Has Been Updated to Process Direct-Infusion Multiple Reaction Monitoring Data. *Lipids* 2018, 53, 1019–1020.
12. Wolrab, D.; Cífková, E.; Čáň, P.; Lísá, M.; Peterka, O.; Chocholoušková, M.; Jirásko, R.; Holčápek, M. LipidQuant 1.0: Automated Data Processing in Lipid Class Separation-Mass Spectrometry Quantitative Workflows. *Bioinformatics* 2021, 37, 4591–4592.

13. Pauling, J.K.; Hermansson, M.; Hartler, J.; Christiansen, K.; Gallego, S.F.; Peng, B.; Ahrends, R.; Ejsing, C.S. Proposal for a Common Nomenclature for Fragment Ions in Mass Spectra of Lipids. *PLoS ONE* 2017, 12, e0188394.
14. Husen, P.; Tarasov, K.; Katafiasz, M.; Sokol, E.; Vogt, J.; Baumgart, J.; Nitsch, R.; Ekroos, K.; Ejsing, C.S. Analysis of Lipid Experiments (ALEX): A Software Framework for Analysis of High-Resolution Shotgun Lipidomics Data. *PLoS ONE* 2013, 8, e79736.
15. Kochen, M.A.; Chambers, M.C.; Holman, J.D.; Nesvizhskii, A.I.; Weintraub, S.T.; Belisle, J.T.; Islam, M.N.; Griss, J.; Tabb, D.L. Greazy: Open-Source Software for Automated Phospholipid Tandem Mass Spectrometry Identification. *Anal. Chem.* 2016, 88, 5733–5741.
16. Kind, T.; Liu, K.-H.; Lee, D.Y.; DeFelice, B.; Meissen, J.K.; Fiehn, O. LipidBlast in Silico Tandem Mass Spectrometry Database for Lipid Identification. *Nat. Methods* 2013, 10, 755–758.
17. Kind, T.; Okazaki, Y.; Saito, K.; Fiehn, O. LipidBlast Templates as Flexible Tools for Creating New In-Silico Tandem Mass Spectral Libraries. *Anal. Chem.* 2014, 86, 11024–11027.
18. Cajka, T.; Fiehn, O. LC-MS-Based Lipidomics and Automated Identification of Lipids Using the LipidBlast In-Silico MS/MS Library. *Methods Mol. Biol.* 2017, 1609, 149–170.
19. Hutchins, P.D.; Russell, J.D.; Coon, J.J. LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. *Cell Syst.* 2018, 6, 621–625.e5.
20. O'Connor, A.; Brasher, C.J.; Slatter, D.A.; Meckelmann, S.W.; Hawksworth, J.I.; Allen, S.M.; O'Donnell, V.B. LipidFinder: A Computational Workflow for Discovery of Lipids Identifies Eicosanoid-Phosphoinositides in Platelets. *JCI Insight* 2017, 2, e91634.
21. Fahy, E.; Alvarez-Jarreta, J.; Brasher, C.J.; Nguyen, A.; Hawksworth, J.I.; Rodrigues, P.; Meckelmann, S.; Allen, S.M.; O'Donnell, V.B. LipidFinder on LIPID MAPS: Peak Filtering, MS Searching and Statistical Analysis for Lipidomics. *Bioinformatics* 2019, 35, 685–687.
22. Alvarez-Jarreta, J.; Rodrigues, P.R.S.; Fahy, E.; O'Connor, A.; Price, A.; Gaud, C.; Andrews, S.; Benton, P.; Siuzdak, G.; Hawksworth, J.I.; et al. LipidFinder 2.0: Advanced Informatics Pipeline for Lipidomics Discovery Applications. *Bioinformatics* 2021, 37, 1478–1479.
23. Ni, Z.; Angelidou, G.; Lange, M.; Hoffmann, R.; Fedorova, M. LipidHunter Identifies Phospholipids by High-Throughput Processing of LC-MS and Shotgun Lipidomics Datasets. *Anal. Chem.* 2017, 89, 8800–8807.
24. Zhou, Z.; Shen, X.; Chen, X.; Tu, J.; Xiong, X.; Zhu, Z.-J. LipidIMMS Analyzer: Integrating Multi-Dimensional Information to Support Lipid Identification in Ion Mobility-Mass Spectrometry Based Lipidomics. *Bioinformatics* 2019, 35, 698–700.
25. Chen, X.; Zhou, Z.; Zhu, Z.-J. The Use of LipidIMMS Analyzer for Lipid Identification in Ion Mobility-Mass Spectrometry-Based Untargeted Lipidomics. *Methods Mol. Biol.* 2020, 2084, 269–282.
26. Koelmel, J.P.; Kroeger, N.M.; Ulmer, C.Z.; Bowden, J.A.; Patterson, R.E.; Cochran, J.A.; Beecher, C.W.W.; Garrett, T.J.; Yost, R.A. LipidMatch: An Automated Workflow for Rule-Based Lipid Identification Using Untargeted High-Resolution Tandem Mass Spectrometry Data. *BMC Bioinform.* 2017, 18, 331.
27. Meng, D.; Zhang, Q.; Gao, X.; Wu, S.; Lin, G. LipidMiner: A Software for Automated Identification and Quantification of Lipids from Multiple Liquid Chromatography-Mass Spectrometry Data Files. *Rapid Commun. Mass Spectrom.* 2014, 28, 981–985.
28. Ahmed, Z.; Mayr, M.; Zeeshan, S.; Dandekar, T.; Mueller, M.J.; Fekete, A. Lipid-Pro: A Computational Lipid Identification Solution for Untargeted Lipidomics on Data-Independent Acquisition Tandem Mass Spectrometry Platforms. *Bioinformatics* 2015, 31, 1150–1153.
29. Herzog, R.; Schuhmann, K.; Schwudke, D.; Sampaio, J.L.; Bornstein, S.R.; Schroeder, M.; Shevchenko, A. LipidXplorer: A Software for Consensual Cross-Platform Lipidomics. *PLoS ONE* 2012, 7, e29851.
30. Herzog, R.; Schwudke, D.; Shevchenko, A. LipidXplorer: Software for Quantitative Shotgun Lipidomics Compatible with Multiple Mass Spectrometry Platforms. *Curr. Protoc. Bioinform.* 2013, 43, 14.12.1–14.12.30.
31. Ross, D.H.; Cho, J.H.; Zhang, R.; Hines, K.M.; Xu, L. LiPydomics: A Python Package for Comprehensive Prediction of Lipid Collision Cross Sections and Retention Times and Analysis of Ion Mobility-Mass Spectrometry-Based Lipidomics Data. *Anal. Chem.* 2020, 92, 14967–14975.
32. Kyle, J.E.; Crowell, K.L.; Casey, C.P.; Fujimoto, G.M.; Kim, S.; Dautel, S.E.; Smith, R.D.; Payne, S.H.; Metz, T.O. LIQUID: An-Open Source Software for Identifying Lipids in LC-MS/MS-Based Lipidomics Data. *Bioinformatics* 2017, 33, 1744–1746.
33. Collins, J.R.; Edwards, B.R.; Fredricks, H.F.; Van Mooy, B.A.S. LOBSTAHS: An Adduct-Based Lipidomics Strategy for Discovery and Identification of Oxidative Stress Biomarkers. *Anal. Chem.* 2016, 88, 7154–7162.
34. Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T.R.; Neumann, S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* 2012, 84, 283–289.
35. Bond, N.J.; Koulman, A.; Griffin, J.L.; Hall, Z. MassPix: An R Package for Annotation and Interpretation of Mass Spectrometry Imaging Data for Lipidomics. *Metabolomics* 2017, 13, 128.

36. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* 2015, 12, 523–526.
37. Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; et al. A Lipidome Atlas in MS-DIAL 4. *Nat. Biotechnol.* 2020, 38, 1159–1163.
38. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinform.* 2010, 11, 395.
39. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* 2006, 78, 779–787.
40. Benton, H.P.; Wong, D.M.; Trauger, S.A.; Siuzdak, G. XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. *Anal. Chem.* 2008, 80, 6382–6389.
41. Peng, B.; Ahrends, R. Adaptation of Skyline for Targeted Lipidomics. *J. Proteome Res.* 2016, 15, 291–301.
42. Peng, B.; Kopczynski, D.; Pratt, B.S.; Ejsing, C.S.; Burla, B.; Hermansson, M.; Benke, P.I.; Tan, S.H.; Chan, M.Y.; Torta, F.; et al. LipidCreator Workbench to Probe the Lipidomic Landscape. *Nat. Commun.* 2020, 11, 2057.
43. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* 2010, 26, 966–968.
44. Ulmer, C.Z.; Koelmel, J.P.; Ragland, J.M.; Garrett, T.J.; Bowden, J.A. LipidPioneer: A Comprehensive User-Generated Exact Mass Template for Lipidomics. *J. Am. Soc. Mass Spectrom.* 2017, 28, 562–565.
45. Song, H.; Hsu, F.-F.; Ladenson, J.; Turk, J. Algorithm for Processing Raw Mass Spectrometric Data to Identify and Quantitate Complex Lipid Molecular Species in Mixtures by Data-Dependent Scanning and Fragment Ion Database Searching. *J. Am. Soc. Mass Spectrom.* 2007, 18, 1848–1858.
46. Goracci, L.; Tortorella, S.; Tiberi, P.; Pellegrino, R.M.; Di Veroli, A.; Valeri, A.; Cruciani, G. Lipostar, a Comprehensive Platform-Neutral Cheminformatics Tool for Lipidomics. *Anal. Chem.* 2017, 89, 6257–6264.
47. Tortorella, S.; Tiberi, P.; Bowman, A.P.; Claes, B.S.R.; Ščupáková, K.; Heeren, R.M.A.; Ellis, S.R.; Cruciani, G. LipostarMSI: Comprehensive, Vendor-Neutral Software for Visualization, Data Analysis, and Automated Molecular Identification in Mass Spectrometry Imaging. *J. Am. Soc. Mass Spectrom.* 2020, 31, 155–163.
48. Kutuzova, S.; Colaianni, P.; Röst, H.; Sachsenberg, T.; Alka, O.; Kohlbacher, O.; Burla, B.; Torta, F.; Schrübbers, L.; Kristensen, M.; et al. SmartPeak Automates Targeted and Quantitative Metabolomics Data Processing. *Anal. Chem.* 2020, 92, 15968–15974.
49. Röst, H.L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; et al. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* 2016, 13, 741–748.
50. Martano, G.; Leone, M.; D'Oro, P.; Matafora, V.; Cattaneo, A.; Masseroli, M.; Bachi, A. SMfinder: Small Molecules Finder for Metabolomics and Lipidomics Analysis. *Anal. Chem.* 2020, 92, 8874–8882.

Retrieved from <https://encyclopedia.pub/entry/history/show/59682>