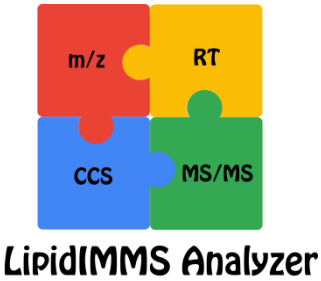


LipidIMMS Analyzer Tutorial

V1.02

Zhiwei Zhou <zhouzw@sioc.ac.cn>

Zheng-Jiang Zhu <jiangzhu@sioc.ac.cn>

August 3, 2018

Laboratory for Mass Spectrometry and Metabolomics

www.zhulab.cn

Interdisciplinary Research Center on Biology and Chemistry (IRCBC)

Shanghai Institute of Organic Chemistry

Chinese Academy of Sciences, Shanghai, China

Table of contents

1. Introduction

Ion mobility - mass spectrometry (IM-MS) has showed great application potential for lipidomics. However, IM-MS based lipidomics is limited by the available tools for lipid identification. **LipidIMMS Analyzer** is developed to support the lipid identification in IM-MS based lipidomics. The software allows to integrate multi-dimensional information including m/z, retention time (RT), collision cross-section (CCS) and MS/MS spectra for lipid identification. Currently, the software supports the IM-MS data acquired from both Agilent and Waters IM-MS instruments, such as Agilent DTIM-MS, Waters Synapt and Vion TWIM-MS.

The software supports different types of data acquisition methods, including:

(1) **LC-IM-MS**: MS1, RT and CCS for lipid identification

(2) **LC-IM-MS/MS** (AIF or MS^E): MS1, RT, CCS and MS/MS spectra for lipid identification

(3) **Direct infusion based IM-MS/MS**: MS1, CCS and MS/MS spectra for lipid identification

The key features of LipidIMMS Analyzer software include:

(1) Provide a large-scale lipid database with four-dimensional structural information;
   a) 4 lipid categories, 25 lipid classes, and 267,716 lipid structures;
   b) Four libraries: MS1 library; RT library; CCS library; and MS/MS spectral library;

(2) Develop an RT calibration method to support LC-system independent application;

(3) Compatible with Agilent Drift tube IM-MS and Waters TMIM-MS techniques;

(4) Integrate multi-dimensional information for lipid identification.

1.1 General workflow

LipidIMMS Analyzer provides an interactive interface for common users to perform lipid identification using lipidomics data acquired from IM-MS. Users could simply import three types of files for the analysis, including **(1) a MS1 peak table; (2) MS/MS spectral data files; and/or (3) an RT calibration table**. Please check **Part 2: Data Preparation** for more instructions.

The workflow includes seven steps:

(1) Data import (a MS1 peak table and/or MS/MS data files);

(2) Select and load lipid database;

(3) Perform the RT calibration (an RT calibration table is required for this step);

(4) Perform m/z, RT and CCS match;

(5) MS/MS spectral match;

(6) Calculate composite score;

(7) Download and browse identification results.

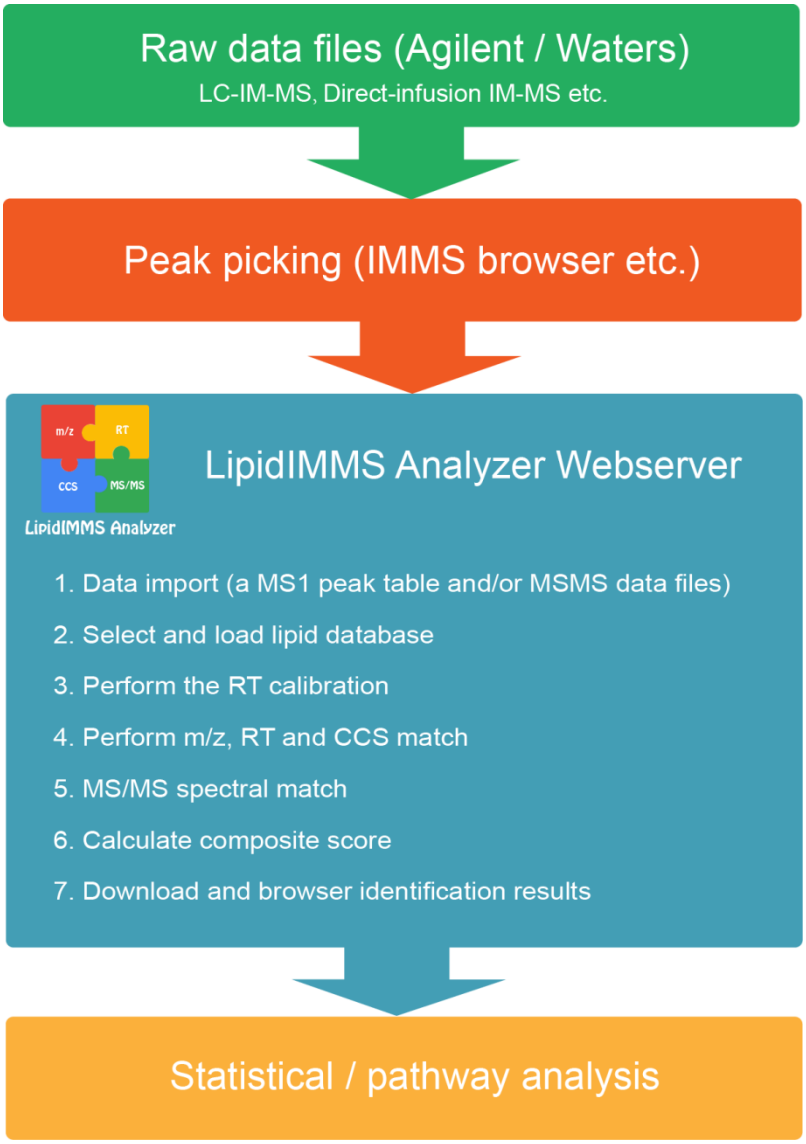


Figure 1.1 The general lipidomics workflow using LipidIMMS Analyzer.

1.2 **Four-dimensional lipid database**

LipidIMMS Analyzer includes a large-scale lipid database with four-dimensional information, including m/z, retention time (RT), collision cross-section (CCS) and MS/MS spectra. As shown in **Table 1**, the database covers **4 categories, 25 classes, and 267,716 lipid structures**. These structures were created using the template-based combinatorial enumeration.[1,2] For each lipid, m/z values for different ion adducts, CCS values, retention times and MS/MS spectra were all calculated and/or predicted using a series of published approaches and/or algorithms.[3,4] CCS values were predicted using our previously published software – LipidCCS.[3] Retention times were predicted using the Random Forest (RF) algorithm. MS/MS spectra were predicted using the fragmentation rules.[2,4] **Finally, a total of 535,432 (267,716 for RP column, and 267,716for HILIC column) RTs, 375,565 CCS values, and 375,565 MS/MS spectra for all lipids in the database were generated.**

Table 1. The information of the lipid database in LipidIMMS Analyzer.

No.	Abbr.	Category	Lipid No.	$[M+H]^+$	$[M+Na]^+$	$[M+NH4]^+$	$[M-H]^-$	$[M+HCOO]^-$
01	PC	GP	8281	8281	8281	0	0	8281
02	pPC	GP	1092	1092	1092	0	0	1092
03	LPC	GP	91	91	91	0	0	91
04	PE	GP	8281	8281	8281	0	8281	0
05	pPE	GP	1092	1092	1092	0	1092	0
06	LPE	GP	91	91	91	0	91	0
07	PS	GP	8281	8281	8281	0	8281	0
08	PG	GP	8281	0	8281	0	8281	0
09	PI	GP	8281	0	8281	0	8281	0
10	PA	GP	8281	0	8281	8281	8281	0
11	SM	SP	1080	1080	1080	0	0	1080
12	ST	SP	168	168	0	0	0	0
13	MG	GL	91	91	0	0	0	0
14	DG	GL	8281	0	0	8281	0	0
15	TG	GL	195,112	0	0	195,112	0	0
16	aLPC	GP	13	13	13	0	0	13
17	pLPE	GP	12	12	12	0	12	0
18	LPS	GP	91	91	0	0	91	0
19	LPG	GP	91	91	0	0	91	0
20	LPI	GP	91	91	91	0	91	0
21	PIP2	GP	8281	8281	8281	0	8281	0
22	LPA	GP	91	0	91	0	91	0
23	Cer	SP	1080	1080	0	0	0	0
24	HexCer	SP	1080	1080	0	0	0	1080
25	Car	FA	102	102	0	0	0	0

All lipids structures are created using the LipidMapsTool[1], expect that Car were imported from LipidBlast[2]. “GP” denotes glycerophospholipids; “GL” denotes glycerolipids; “SP” denotes sphingolipids; “FA” denotes fatty acids.

Table 2. The abbreviation table of lipid classes.

No.	Abbreviation	Name
1	PC	Glycerophosphatidylcholine
2	pPC	Plasmenyl-glycerophosphatidylcholine
3	LPC	Lysoglycerophosphatidylcholine
4	PE	Glycerophosphatidylethanolamine
5	pPE	Plasmenyl-glycerophosphatidylethanolamine
6	LPE	Lysoglycerophosphatidylethanolamine
7	PS	Glycerophosphatidylserine
8	PG	Glycerophosphatidylglycerol
9	PI	Glycerophosphatidylinositol
10	PA	Glycerophosphatidic acid
11	SM	Sphingomyelin
12	ST	Sulfatide
13	MG	Monoacylglycerol
14	DG	Diacylglycerol
15	TG	Triacylglycerol
16	aLPC	Plasmanyl-lysoglycerophosphatidylcholine
17	pLPE	Plasmenyl-lysoglycerophosphatidylethanolamine
18	LPS	Lysoglycerophosphatidylserine
19	LPG	Lysoglycerophosphatidylglycerol
20	LPI	Lysoglycerophosphatidylinositol
21	PIP2	Glycerophosphoinositol bisphosphate
22	LPA	Lysoglycerophosphatidic acid
23	Cer	Ceramide
24	HexCer	Hexosylceramide
25	Car	Carnitine

1.3 **Retention time calibration**

To apply the RT match in different LC conditions and correct RT drift in different experiments, an RT calibration method was developed using a list of 20 lipids to re-calibrate all RT values in the database (**Table 3-4**). Users could select some of these lipids for the calibration. For convenience, users could directly use a commercial available lipid mixture (Differential Ion Mobility System Suitability Lipidomix, DIMS Kit, Avanti Lipids), which contains 10 different lipids in the calibration list. The basic principle of RT calibration is that the elution order of lipids is conserved for similar LC systems.[5] First, a Locally Weighted Scatterplot Smoothing (LOESS) calibration model is built between the experiment RTs and reference RTs of lipid standards. Then RTs in the database are re-calculated to match the experimental condition.

Table 3. The retention times of lipids for RT calibration on a reverse phase column.

No.	Lipid name	Adduct^{pos}	m/z^{pos}	Adduct^{neg}	m/z^{neg}	RT(s)
1	TG(18:1/18:1/18:1)	$[M+NH_4]^+$	902.8171	n.a.	n.a.	659
2	SM(d18:1/18:1)	$[M+H]^+$	729.5905	$[M+HCOO]^-$	773.5809	332
3	Cer(d18:1/18:1)	$[M+H]^+$	564.5350	$[M+HCOO]^-$	608.5254	384
4	LPC(18:1)	$[M+H]^+$	522.3554	$[M+HCOO]^-$	566.3458	86
5	DG(14:1/14:1)	$[M+NH_4]^+$	526.4472	n.a.	n.a.	283
6	PC(14:1/14:1)	$[M+H]^+$	674.4755	$[M+HCOO]^-$	718.466	193
7	PS(14:1/14:1)	$[M+H]^+$	676.4184	$[M-H]^-$	674.4099	142
8	PG(14:1/14:1)	$[M+Na]^+$	685.4051	$[M-H]^-$	661.4081	135
9	PE(14:1/14:1)	$[M+H]^+$	632.4286	$[M-H]^-$	630.4135	190
10	PI(14:1/14:1)	$[M+Na]^+$	773.4211	$[M-H]^-$	749.4241	122
11	TG(14:1/14:1/14:1)	$[M+NH4]^+$	734.6299	n.a.	n.a.	516
12	SM(d18:1/12:0)	$[M+H]^+$	647.5123	$[M+HCOO]^-$	691.5027	218
13	Cer(d18:1/12:0)	$[M+H]^+$	482.4568	$[M+HCOO]^-$	526.4472	254
14	LPC(24:0)	$[M+H]^+$	608.4650	$[M+HCOO]^-$	652.4554	201
15	DG(17:1/17:1)	$[M+NH4]^+$	610.5411	n.a.	n.a.	415
16	PC(17:0/17:0)	$[M+H]^+$	762.6008	$[M+HCOO]^-$	806.5912	394
17	PS(17:0/17:0)	[M+H]+	764.5436	$[M-H]^-$	762.5291	344
18	PG(17:0/17:0)	$[M+Na]^+$	773.5303	$[M-H]^-$	749.5338	363
19	PE(17:0/17:0)	$[M+H]^+$	720.5538	$[M-H]^-$	718.5392	382
20	PI(16:0/18:1)	$[M+Na]^+$	859.5307	$[M-H]^-$	835.5342	305

Note: **retention times of number 1-10 lipids were experimentally acquired from DIMS Kit**, and the retention times of 11-20 lipids were predicted. the superscripts “pos” and “neg” represent positive and negative modes, respectively; “n.a.” refers to “not available”.

Table 4. The retention times of lipids for RT calibration on a HILIC column.

No.	Lipid name	Adduct^{pos}	m/z^{pos}	Adduct^{neg}	m/z^{neg}	RT(s)
1	TG(18:1/18:1/18:1)	$[M+NH_4]^+$	902.8171	n.a.	n.a.	159
2	SM(d18:1/18:1)	$[M+H]^+$	729.5905	$[M+CH3COO]^-$	787.5971	434
3	Cer(d18:1/18:1)	$[M+H]^+$	564.5350	$[M-H]^-$	562.5205	32
4	LPC(18:1)	$[M+H]^+$	522.3554	$[M+CH3COO]^-$	580.3620	470
5	DG(14:1/14:1)	$[M+NH_4]^+$	526.4472	n.a.	n.a.	38
6	PC(14:1/14:1)	$[M+H]^+$	674.4755	$[M+CH3COO]^-$	732.4821	404
7	PS(14:1/14:1)	$[M+H]^+$	676.4184	$[M-H]^-$	674.4099	366
8	PG(14:1/14:1)	$[M+Na]^+$	685.4051	$[M-H]^-$	661.4081	150
9	PE(14:1/14:1)	$[M+H]^+$	632.4286	$[M-H]^-$	630.4135	359
10	PI(14:1/14:1)	$[M+Na]^+$	773.4211	$[M-H]^-$	749.4241	223
11	TG(14:1/14:1/14:1)	$[M+NH4]^+$	734.6299	n.a.	n.a.	157
12	SM(d18:1/12:0)	$[M+H]^+$	647.5123	$[M+CH3COO]^-$	705.5188	436
13	Cer(d18:1/12:0)	$[M+H]^+$	482.4568	$[M-H]^-$	480.4422	36
14	LPC(24:0)	$[M+H]^+$	608.4650	$[M+CH3COO]^-$	666.4716	461
15	DG(17:1/17:1)	$[M+NH4]^+$	610.5411	n.a.	n.a.	34
16	PC(17:0/17:0)	$[M+H]^+$	762.6008	$[M+CH3COO]^-$	820.6073	402
17	PS(17:0/17:0)	[M+H]+	764.5436	$[M-H]^-$	762.5291	353
18	PG(17:0/17:0)	$[M+Na]^+$	773.5303	$[M-H]^-$	749.5338	123
19	PE(17:0/17:0)	$[M+H]^+$	720.5538	$[M-H]^-$	718.5392	331
20	PI(16:0/18:1)	$[M+Na]^+$	859.5307	$[M-H]^-$	835.5342	223

Note: retention times of number 1-20 lipids were predicted using the data provided in literature[6]; the superscripts “pos” and “neg” represent positive and negative modes, respectively; “n.a.” refers to “not available”.

1.4 **Scoring system for multi-dimensional lipid identification**

LipidIMMS Analyzer enables to integrate multi-dimensional information including *m/z*, RT, CCS, and MS/MS spectra for lipid identification in IM-MS. First, the software performed accurate mass match using the user-defined *m/z* tolerance (e.g., 25 ppm). Then, each lipid candidate was further evaluated and scored through comparing their experimental RT, CCS, and MS/MS spectra to those in the database. RT and CCS matches were scored using a trapezoidal function (**Figure 1.2a**). MS/MS spectral match was scored using a reverse dot-product function (**Figure 1.2b**). Finally, the composite score was calculated using a linear weighting function according to the user-defined weight for each match score (**Figure 1.2c**).

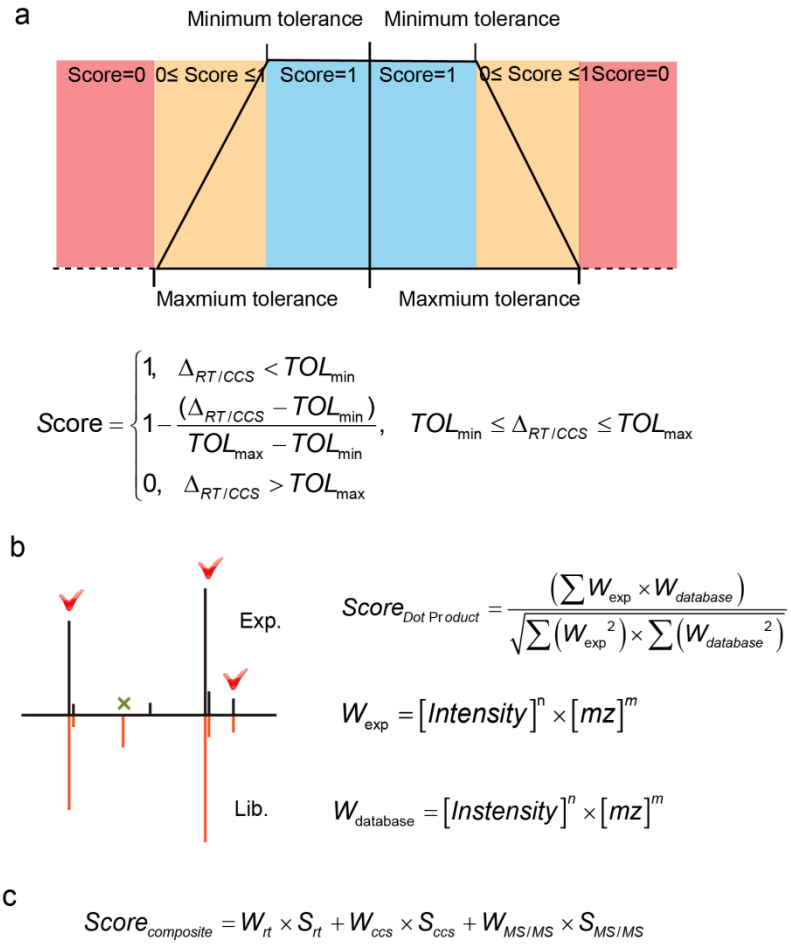


Figure 1.2 The scoring functions in LipidIMMS Analyzer. (**a**) The trapezoidal function for scoring RT and CCS matches; (**b**) the reverse dot-product function for scoring MS/MS spectral match. The weights for intensity (n) and *m/z* (m) were set as: n=0.6, m=1 for positive mode, and n=1, m=1 for negative mode, respectively; (**c**) the equation to calculate the composite score by integrating multiple match scores. The abbreviation "TOL_{max}" denotes "maximum tolerance" and the abbreviation "TOL_{min}" denotes "minimum tolerance".

2. Data Preparation

2.1 Overview

LipidIMMS Analyzer supports both Agilent and Waters instruments. It requires three types of files for the analysis **(Figure 2.1)**:

(1) A MS1 peak table (.csv format)

(2) MS/MS data files (.mgf/.msp /.cef format)

(3) An RT calibration table (.csv format, only required for RT match).

Please download "demo data" in "Help" tab in the webserver.

The MS1 peak table is a list of peaks with m/z, retention time (RT), collision cross-section (CCS) and peak intensities. The MS2 data files are the MS/MS spectra for MS1 peaks. The RT calibration table is a table used for RT calibration. We recommend the users to use vendors' software for peak picking and generating the MS1 table peak and MS/MS data files.

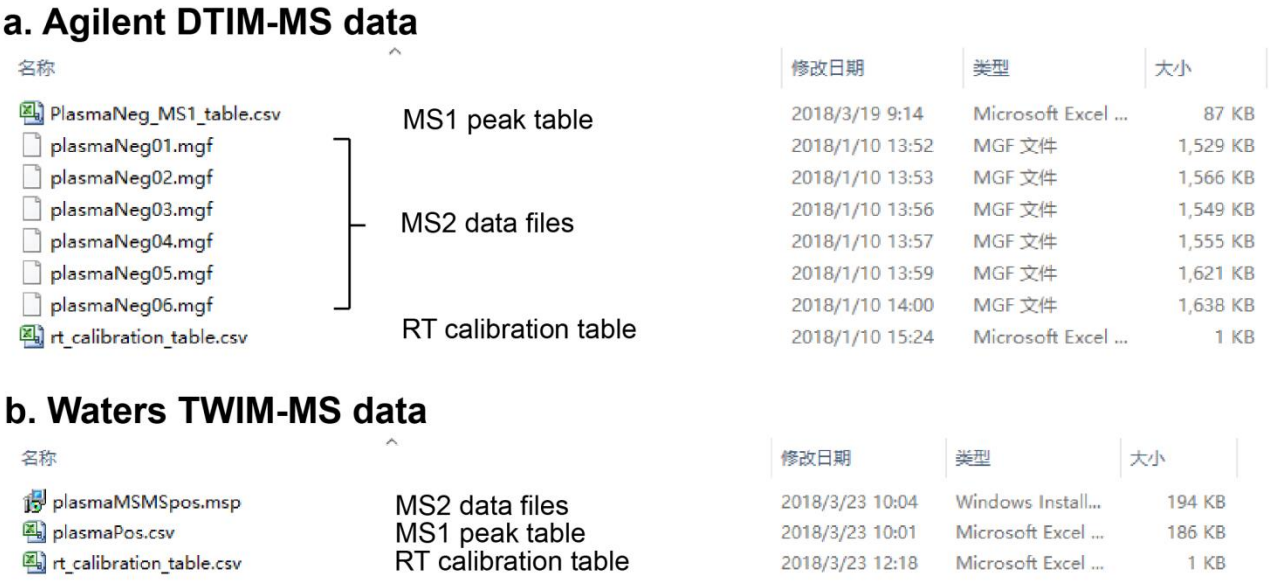


Figure 2.1 The imported data files for the LipidIMMS Analyzer.

In the following sections, we will provide a **step-by-step instruction** to prepare the required files from raw data files.

2.2 Agilent DTIM-MS

Required tools:

- Agilent Mass Profiler (Version B.08.00 or later)
- Agilent IM-MS Browser (Version B.08.00 or later)

(1) Raw data processing (Single-field based data acquisition)

a. Post-calibrate raw data in **IM-MS Reprocessor** software from Agilent.
b. Calculate the calibration coefficient (T_{fix} and β) of calibrants (Agilent tune mixture solution) using **Agilent IM-MS Browser**, and save calibration coefficient to all data files.
c. Open files in **Agilent Mass Profiler**, and select the appropriate parameters to peak detection and alignment. In the demo data, we use the following parameters:

Feature Finding/Loading:

- Measure of abundance-Max ion volume;
- Ion intensity >= 100 count;
- Isotope model: Common organic molecules
- Limit charge states to a range of 1-1

Alignment & Normalization:

- RT tolerance = ±(0.0% + 0.3 min)
- Mass tolerance = ±(15ppm + 0.2 mDa)

Statistics & Filters:

- Missing sample treatment: Assign 0 abundance
- Feature filter: Q-score >= 70.0
- Sample occurrence: Frequency >= 50% in at least one group

(2) Export the MS1 peak table

Export the MS1 peak table in Agilent Mass Profiler, and modify the peak table as the following format:

(1) The first column is set as the mass-to-charge ratio ("mz").
(2) The second column is set as the retention time ("rt").
(3) The third column is set as the collision cross-section ("ccs").
(4) Other columns are set as peak abundances in each sample.

IMPORTANT NOTES:

1. The order and names of the first three columns must be "mz", "rt" and “ccs”.
2. The unit of retention time must be minute.
3. The csv file must be separated by comma.

(3) Export MS/MS data files (MGF)

After the peak detection using Mass Profiler, open data files using **Agilent IM-MS browser**. Then, export MS/MS spectra of all features in Mascot Generic Format (*.mgf). We highly suggest only using MS/MS data files form the pooled quality control samples.

IMPORTANT NOTES:

1. The MS/MS data files have to be exported one by one due to the limitation of IM-MS browser software. An update will be provided by Agilent to export all MS/MS data files all together.
2. We recommend exporting the top 100 fragment ions for each feature. Please set this parameter in the “Method - Find peaks in mass spectrum – Maximum peak count”.
3. Only **Mascot Generic Format (*.mgf)** is supported for Agilent DTIM-MS data.

(4) Export MS/MS data files (CEF)

For the version of **Agilent Mass Profiler B.08.01 (B153, Beat Release)**, it supports to directly export MS/MS spectra for each detected feature as CEF format instead of using Agilent IM-MS Browser. After the raw data processing, users could export MS/MS spectra in the tab “File”, then select “Export Each Sample to CEF”. Please refer the **Figure 2.2** for procedures and export parameters.

IMPORTANT NOTES:

1. This function was only available for **Mass Profiler B.08.01 (B153, Beat Release) or latter**.
2. It was recommended to put an additional parameter file “*FragmentDriftTimeOffset.txt*” **(Figure 2.3)** in the “*C:\temp*” folder of your computer to correct the small shift of drift time between the precursor and fragment ions.

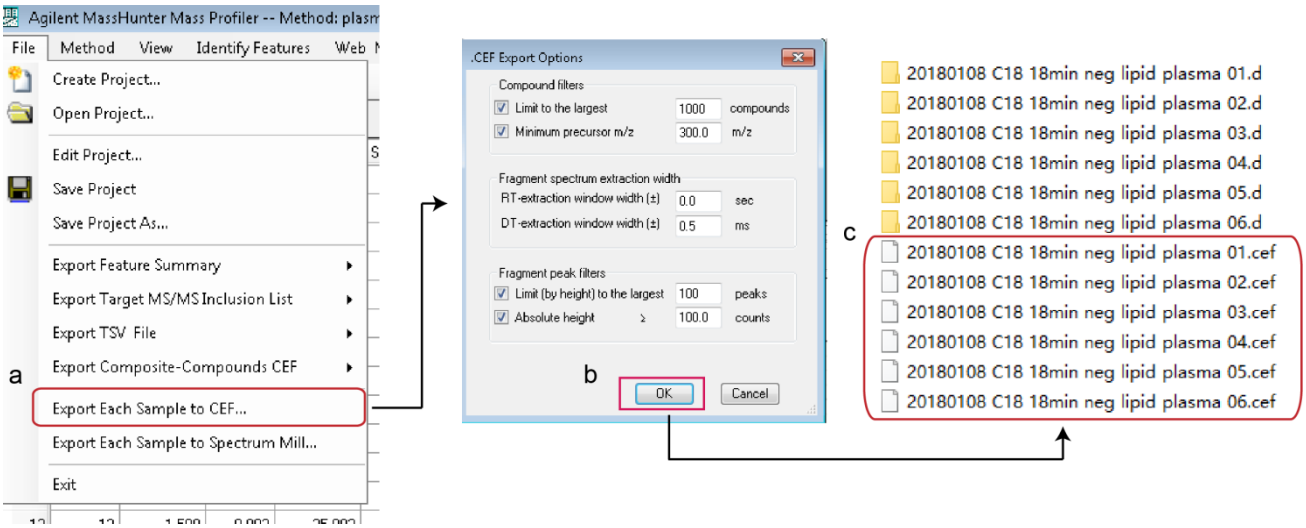


Figure 2.2 The procedures for exporting the MS/MS spectra file in CEF format using Mass Profiler.

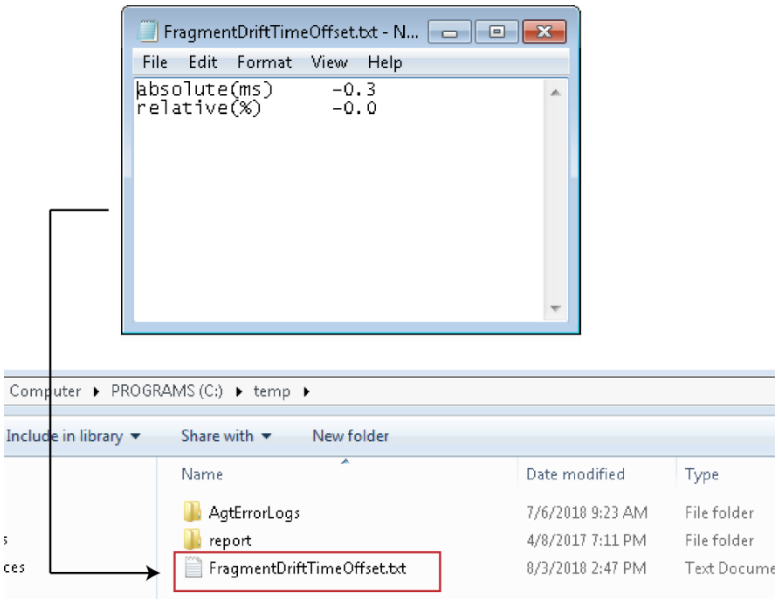
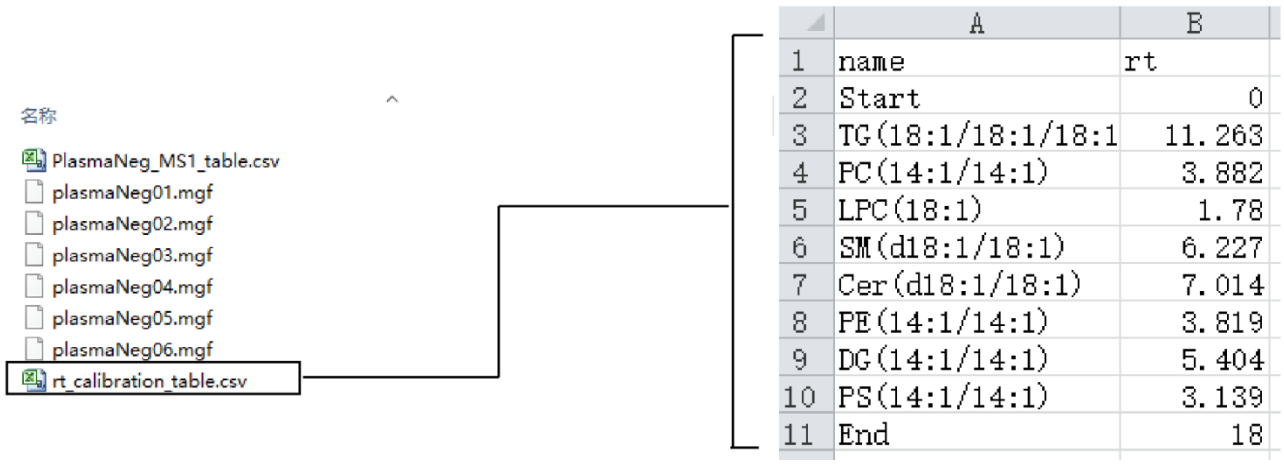


Figure 2.3 The addition of a "FragmentDriftTimeOffset.txt" file in "C:\temp" folder.

(5) Prepare the RT calibration table

The RT calibration table is prepared in a specific format (.csv file). The first column is "name" of the lipids in DMS Kit (Please see selection 1.3). The second column "rt" is the experimental retention times of lipids. An example of RT calibration table is given below:



	A	B
1	name	rt
2	Start	0
3	TG(18:1/18:1/18:1	11.263
4	PC(14:1/14:1)	3.882
5	LPC(18:1)	1.78
6	SM(d18:1/18:1)	6.227
7	Cer(d18:1/18:1)	7.014
8	PE(14:1/14:1)	3.819
9	DG(14:1/14:1)	5.404
10	PS(14:1/14:1)	3.139
11	End	18

Figure 2.4 The screen shot of the imported RT calibration table.

IMPORTANT NOTES:

1. The **"Start"** and **"End"** represents the start and end points of gradient. It must be filled.
2. The name of lipids must be consistent with the template.
3. If some lipids were not detected in your system, please remove the rows of missing lipids.
4. The unit of RT must be **minute**.

2.3 Waters TWIM-MS

Required tools:

- Progenesis QI (Version 2.3 or later)

(1) Raw data processing

Please refer to **Progensis QI tutorial** for the raw data processing. Some critical parameters of data processing used in our demo data set were as following:

- Alignment reference: Assess all runs in the experiment for suitability
- Sensitivity: Automatic
- Adducts: M+H-H_2O, M+H, M+NH_4, M+Na

IMPORTANT NOTE:

1. The accuracy of CCS values from TWIM-MS depends on the structural similarity between the analytes and the calibrant ions. It is highly recommended to use the **lipid calibrants** for the calculation of CCS values.[6]

(2) Export the MS1 peak table

Users could export the MS1 peak table in the step of “Review Compounds”. The exported .csv table is required to be modified as the following format:

(1) The first column is set as the mass-to-charge ratio ("mz").
(2) The second column is set as the retention time ("rt").
(3) The third column is set as the collision cross-section ("ccs").
(4) The fourth column is the compound name provided by Progenesis QI ("compound").
(5) Other columns are peak abundances of MS1 peaks in each sample.

IMPORTANT NOTES:

1. The order and names of the first four columns must be "mz", "rt", “ccs” and “compound”.
2. The unit of retention time must be minute.
3. The csv file must be separated by comma.
4. **The only difference of the MS1 peak tables between Agilent and Waters is the “compound” column.**

(3) Export of MS/MS spectrum

The MS/MS spectra in Progensis QI were exported in the “Review Compounds - Export fragment database”. The MS/MS spectra are saved as the format of “Mass Spectral Database (*.msp)”. Only **Mass Spectral Database (*.msp)** is supported for Waters TWIM-MS data.

(4) Prepare the RT calibration table

The requirement of the RT calibration table for Waters data is same as Agilent data (Section 2.3.4).

3. The use of LipidIMMS Analyzer

3.1 The layout of LipidIMMS Analyzer

The LipidIMMS Analyzer provides an interactive interface to help users to analyze the data from IM-MS. It consists of three parts for each page (**Figure 3.1**), including: “Stage navigation panel”, “Parameter setting panel”, and “Result display panel”. Users just click the first choice in “stage navigation panel” to start the analysis. In each stage, users could set parameters in the “parameter setting panel”, and results will be quickly returned in the “result display panel”.

- **Stage navigation panel:** providing the stage navigation
- **Parameter setting panel:** setting analysis parameters
- **Result display panel:** displaying analysis result

In the following sections, we will demonstrate how to use the LipidIMMS Analyzer step by step.

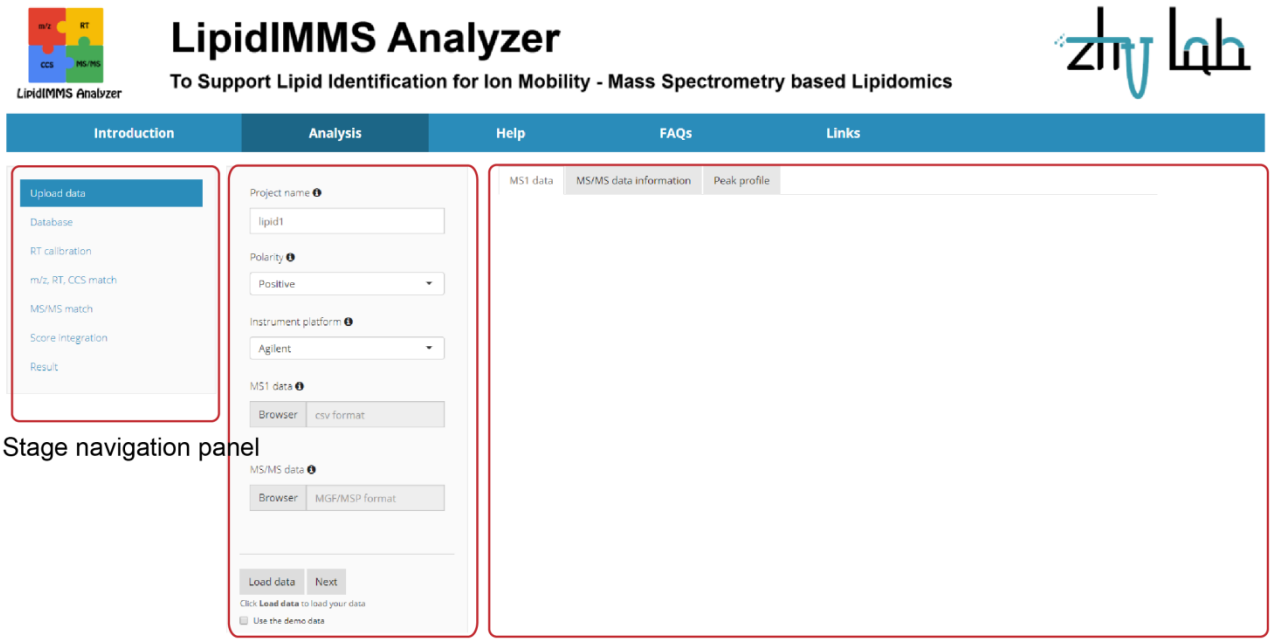


Figure 3.1 The layout of LipidIMMS Analyzer webserver.

3.2 Use of LipidIMMS Analyzer

Step 1. Upload data

a) Input basic information of the project, including “Project name”, “Polarity”, “Instrument platform”.
b) Upload MS1 data (MS1 peak table) and MS/MS data.
c) Click the “Load data”.
d) View results in the “MS1 data”, “MS/MS data information” and “Peak profile” tabs. “MS1 data” displays the uploaded MS1 peak table; “MS/MS data information” displays the statistical information of the uploaded MS/MS data files; “Peak profile” displays the plot of peak profile, including “m/z vs. RT” and “m/z vs. CCS” plots.
e) Click the “Next” to proceed to next step.

Note: If no MS/MS data files available, one can only upload the MS1 data.

Parameter definition:

- Project name: Required. The name of the project.
- Polarity: Required. The ionization polarity.
- Instrument platform: Required. The instrument platform.
- MS1 data: Required. A csv file is with a specific format.
- MS/MS data: Optional. MGF or MSP files with a total size up to 200M.

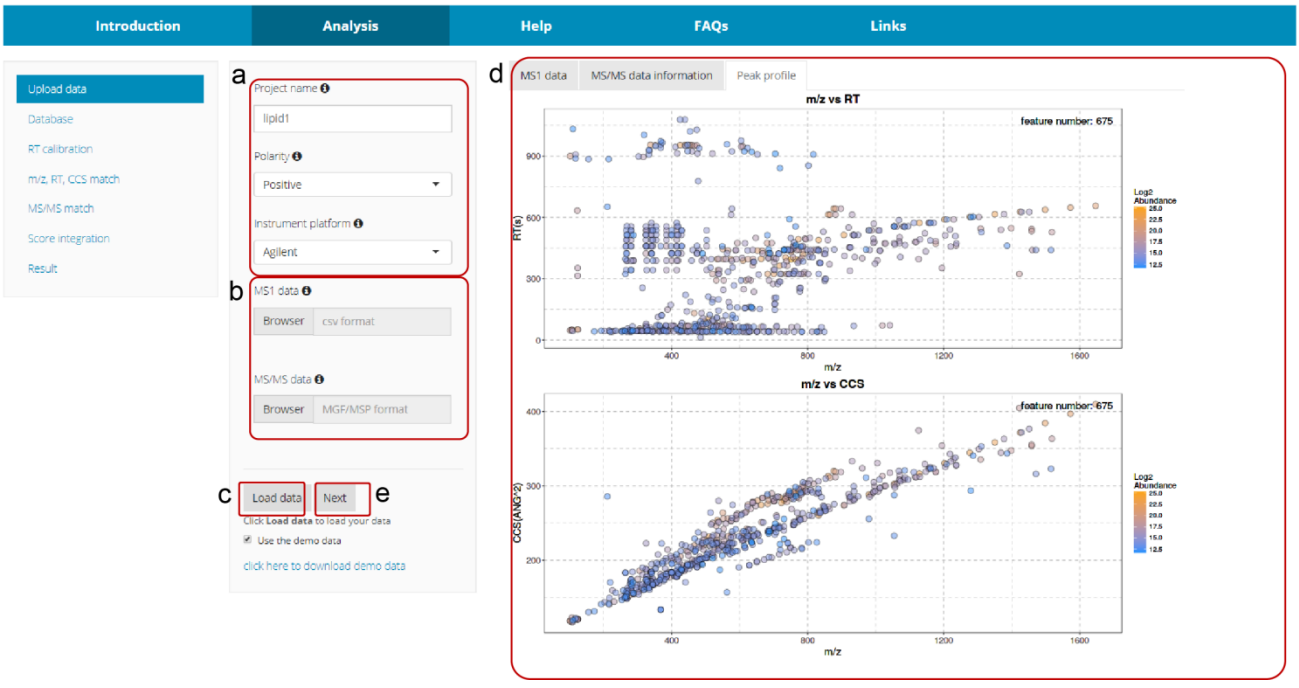


Figure 3.2 The interface of the “Upload data” tab.

Step 2. Database

a) Select specific lipid classes for identification. All lipid classes are chosen by default.
b) Click the “Submit”.
c) View and check the lipid classes in “Database information” tab. The abbreviations of lipid classes are listed in the “Abbreviations” tab.
d) Click the “Next” to proceed to next step.

Parameter definition:

- Included: Selected lipid classes for lipid identification.
- Excluded: Excluded lipid classes from lipid identification.

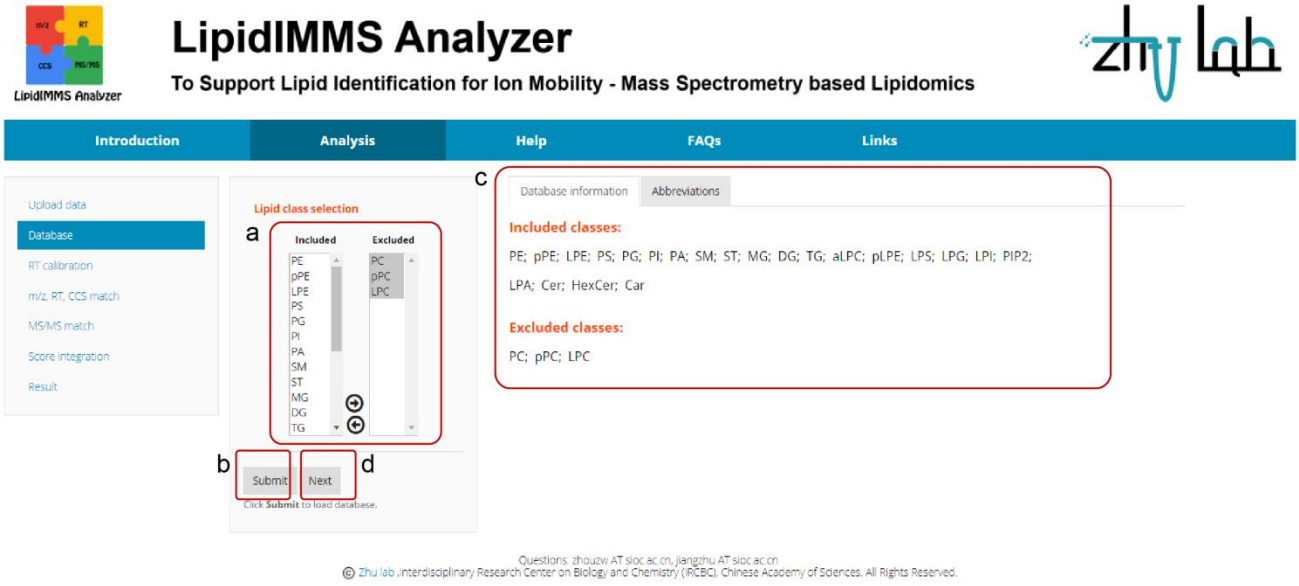


Figure 3.3 The interface of the “Database” tab.

Step 3. RT calibration

a) Select “Column type”, and upload the RT calibration table.
b) Click the “Submit”.
c) View and check the uploaded table in “RT calibration table” tab. The “RT calibration plot” tab display the fitted curve of calibration.
d) Click the “Next” to proceed to next step.

Note:

- If users want to skip the RT calibration, please directly click the “Next”.
- RT calibration is required for the following RT match.

Parameter definition:

- Column type: Required. The column type for LC separation.
- RT calibration table: Optional. A csv table is required in a specific format, which a mandatory file to perform RT match. If you do NOT perform RT match, please click next button.

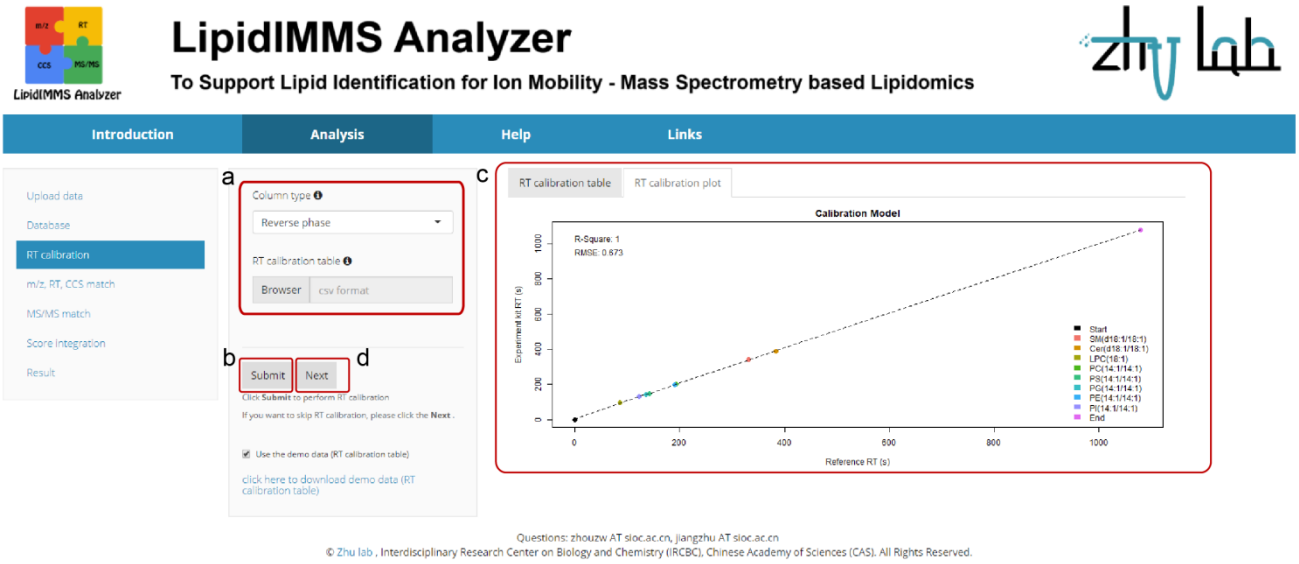


Figure 3.4 The interface of “RT calibration” tab.

Step 4. m/z, RT, CCS match

a) Input m/z tolerance for match.
b) Check or uncheck “RT match”; Input the minimum and maximum tolerances for RT match score.
c) Check or uncheck “CCS match”; Input the minimum and maximum tolerances for CCS match score.
d) Click the “Submit”.
e) View match result in “Match result” tab. In the table, lipid candidates will be displayed both in “lipid species” and “lipid molecular species” (**See Section 4**). In addition, m/z error, rt error, rt score, CCS error, and CCS score will also be displayed. Only top 10 candidates for each peak are given in the result table. One can select “All” entries, and download the result table in either .csv or Excel format.
f) Click the “Next” to proceed to next step.

Note:

- Accurate mass match is required.
- If you want to skip the “RT match” and/or “CCS match”, please uncheck the “RT match” and/or “CCS match”.
- If the “RT match” and/or “CCS match” is unchecked, the inputted parameters would be invalid.

Parameter definition:

Accurate Mass

- m/z tolerance (ppm): Required. Range: 0-500 ppm.

Retention time

- Minimum tolerance (s): Required. If error is within the tolerance, RT match score equals to 1. Range: 0-300 s.
- Maximum tolerance (s): Required. If error is larger the tolerance, RT match score equals to 0, and lipid candidates will be removed. Range: 0-300 s.
- RT match: Optional. If this is not checked, parameters of retention time match are invalid.

Collision Cross Section

- Minimum tolerance (s): Required. If error is within the tolerance, CCS match score equals to 1. Range: 0-100%.
- Maximum tolerance (s): Required. If error is larger than the tolerance, CCS match score equals to 0, and lipid candidates will be removed. Range: 0-100%.
- CCS match: Optional. If this is not checked, parameters of CCS match are invalid.

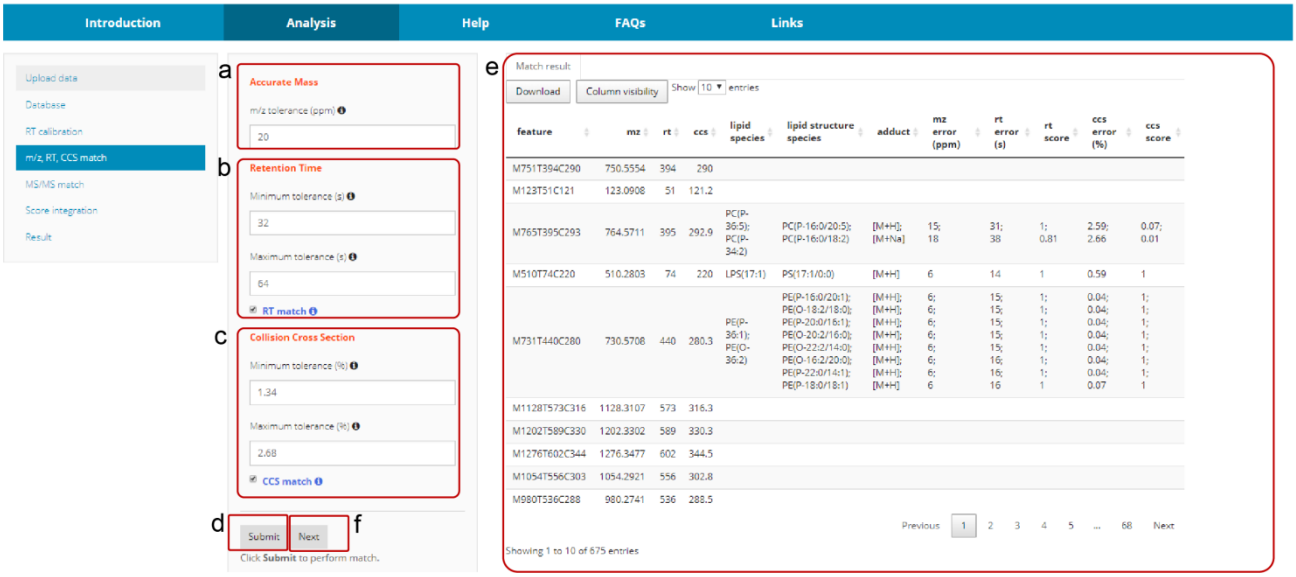


Figure 3.5 The interface of “m/z, RT, CCS match” tab.

Step 5. MS/MS match

a) Check or uncheck the “MS/MS match”.
b) Set the mass range for MS/MS spectra.
c) Input the values of “absolute intensity cutoff”, “relative intensity cutoff” and “MS/MS score cutoff”.
d) Click the “Submit”.
e) View match result in “Match result” tab. In this table, lipid candidates are both displayed in “lipid species” and “lipid molecular species”. MS/MS spectral match scores (“msms score”) are calculated. Please refer to **Section 4** for more details about the result.
f) Click the “Next” to proceed to next step.

Note:

- If you want to skip the “MS/MS match”, please uncheck the “MS/MS match”.
- If uncheck the “MS/MS match”, the inputted parameters would be invalid.

Parameter definition:

- MS/MS mass range: Required. Only MS2 fragments within the mass range are reserved for spectral match. Range: 0-2000 Da.
- Absolute intensity cutoff: Required. If the fragment intensities are smaller than the cutoff value, the fragments will be removed from the MS/MS spectra. Range: 0-2000 counts.
- Relative intensity cutoff: Required. If intensity ratios of fragments compared to the base peak are smaller than the cutoff value, the fragments will be removed from the MS/MS spectra. Range: 0-1.
- MS/MS score cutoff: Required. If the match score is smaller than the cutoff score, the lipid candidate will be removed. Range: 0-1.
- MS/MS match: Optional. If this is not checked, the parameters of MS/MS match will be invalid.
- Advanced parameter: Optional. Some parameters for MS/MS spectra selection. We strongly recommend the default parameters here!

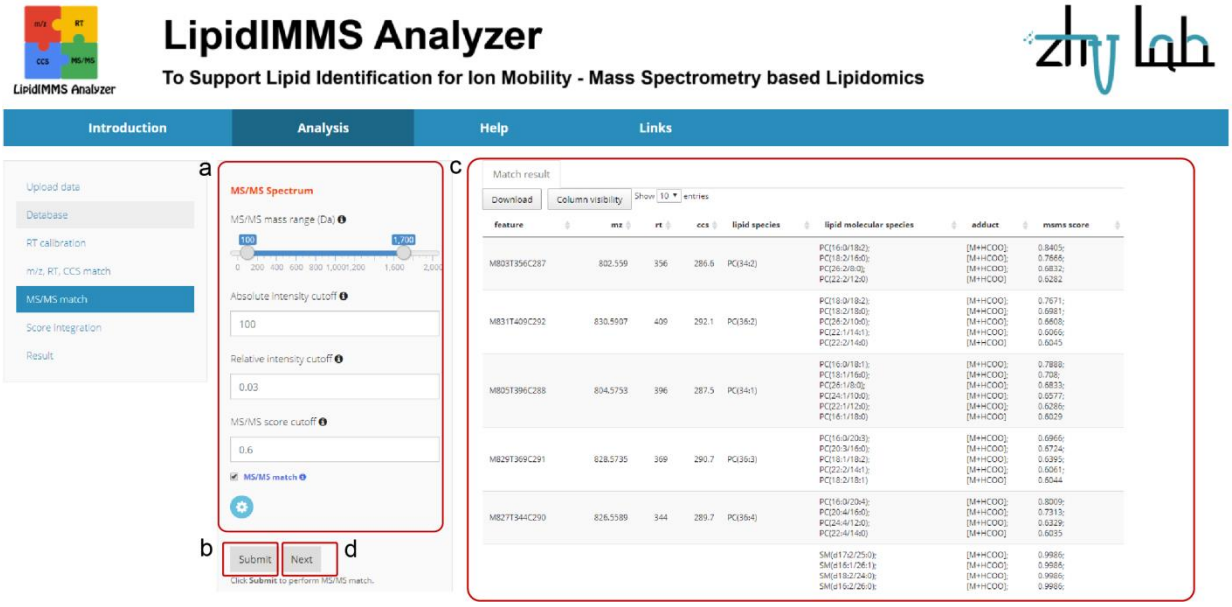


Figure 3.6 The interface of “MS/MS match” tab.

Step 6. Score Integration

a) Input the values of “RT score weight”, “CCS score weight”, and “MS/MS score weight”, respectively, to calculate the composite score; Input the value of “Composite score cutoff”.

b) Click the “Submit”.

c) View match results in “Match result” tab. In this table, composite score is calculated, and lipid candidates are given if the score is larger than the cutoff score.

d) Click the “Next” to proceed to next step.

Note:

- The sum of three weights should be equal to 1.
- If any match is unchecked, the corresponding score weight should be 0. For example, if the RT match is unchecked, the RT score weight should be 0.
- If no match was performed, please set all weights to 0. Then, click the “Submit”.

Parameter definition:

- RT score weight: Required. The weight of RT score is used to calculate the composite score. Range: 0-1.
- CCS score weight: Required. The weight of CCS score is used to calculate the composite score. Range: 0-1.
- MS/MS score weight: Required. The weight of MS/MS score is used to calculate the composite score. Range: 0-1.
- Composite score cutoff: Required. If composite score is less than the cutoff value, lipid candidates are removed. Range: 0-1.

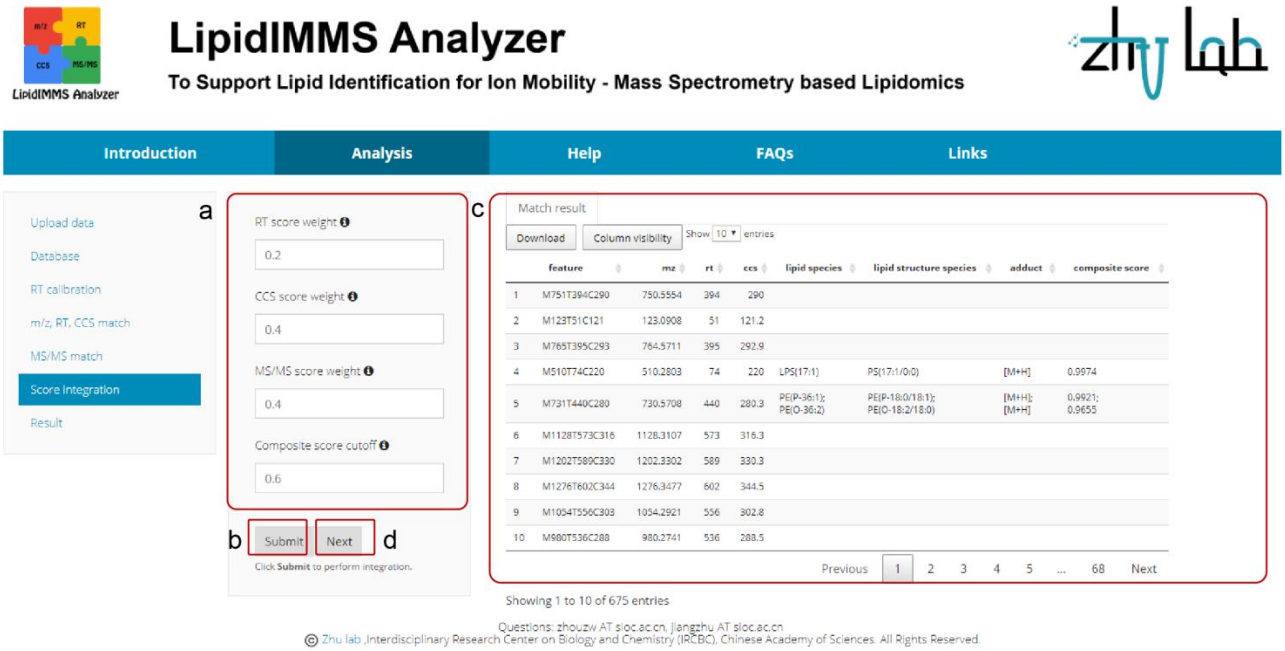


Figure 3.7 The interface of “Score integration” tab.

Step 7. Result

In result page, it includes three tabs: “Summary”, “Result table” and “Detail”.

Summary: It summarizes the analysis result, and users could download a ZIP file of analysis result. The zip file includes an html report and an identification table.

Result table: Lipid identifications for each feature are listed in the result table, and user could browser them in the webserver. All identified lipids are displayed as both “lipid species” and “lipid molecular species” levels. Please refer to **Section 4** for the definition of these levels. For each feature, the table only displays the top 10 lipid candidates ranked by their composite scores. Users could select a row of interested feature, and view all identification information in the “Detail” tab.

Detail: This tab displays all candidates for the user selected feature. It contains all identification information, including m/z error, CCS error, MS/MS score, and composite score and so on. If users want to visualize the **MS/MS spectral match plot**, they could select the specific row of one candidate, and click the **“show MS/MS plot”**.

We will interpret the details of the analysis result in the **Section 4**.

4. Interpretation of Analysis Result

The result zip file could be downloaded in the **"Summary" tab** in the final "Result" step. The downloaded zip file contains 2 files: "analysis report" and "result table".

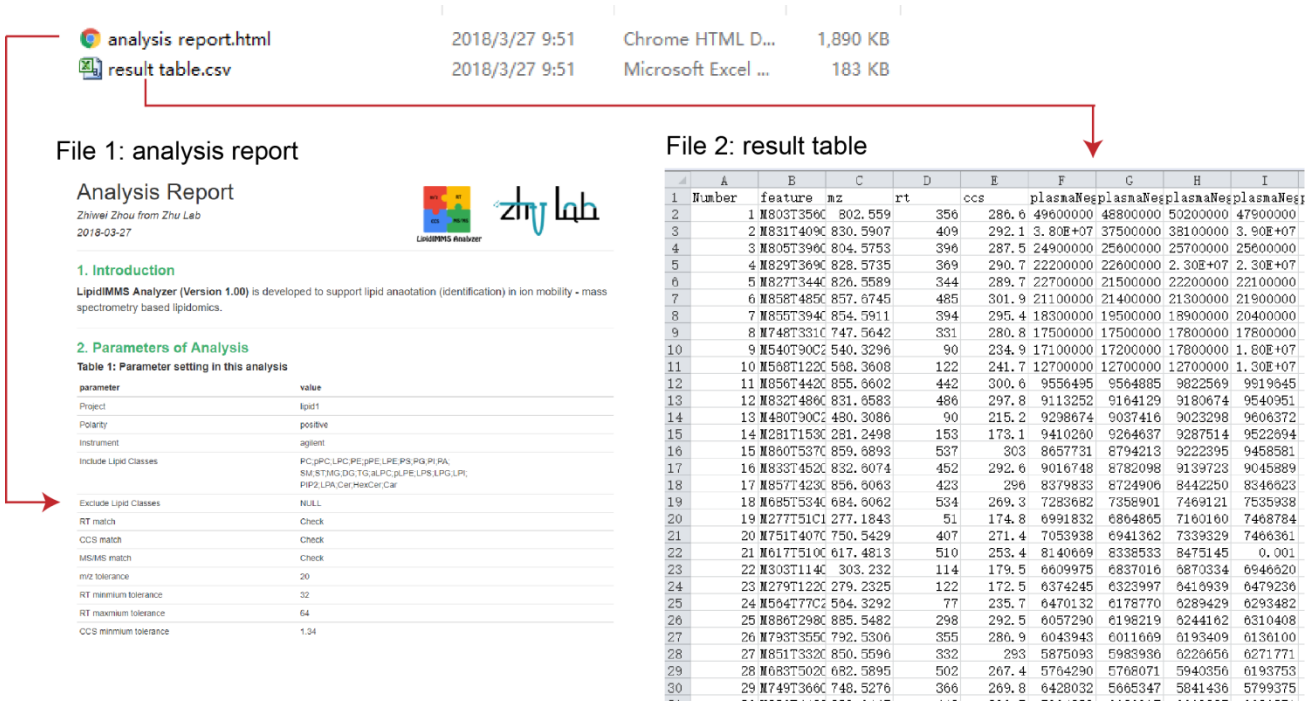


Figure 4.1 Schematic illustration of the result zip file.

1. **Analysis report:** This report is created in the html format, and users could open it using common browsers like Chrome, Firefox, Safari etc. It includes four components: "Introduction", "Parameters of Analysis", "Result", and "Plots".

- **Introduction:** It records the version of LipidIMMS Analyzer in the analysis.
- **Parameters of Analysis:** It displays a table to record the parameters of the analysis.
- **Result:** It summarizes the total numbers of identifications, feature numbers and lipid candidates in the analysis.
- **Plots:** It displays several figures in the analysis, including peak profile, and RT calibration plot.

2. **Result table:** The result table is generated in a csv format. It includes all identification results in the MS1 peak table. The explanation of the columns are listed as followings:

- **Number:** The feature or peak number.
- **Feature:** The name of each peak, it defined by m/z, RT and CCS. For example, M803T356C287 represents a feature with m/z 802.559, RT 356 s, and CCS 286.6 $Å^2$.

- **Lipid species:** The lipid candidates defined in “lipid species” level and the order is ranked by their composite scores. The definition of ‘lipid species’ could see reference 7. It refers to lipid subclass, and characterizes the lipids by the number of carbons and double bonds. For example, PC(36:2) represents that a glycerophosphatidylcholine has 36 carbons and 2 double bonds.
- **Lipid molecular species:** The lipid candidates defined in “lipid molecular species” level, and the order is ranked by their composite scores. The definition of ‘lipid molecular species’ could see reference 7. It characterizes the lipids in fatty acyl position level. For example, PC(20:4/16:0) represents that it has two acyl chains with 20:4 and 16:0 fatty acyls in in sn-1 and sn-2, respectively.
- **Adducts:** The adduct forms of lipid candidates. The adduct forms of each lipid class is listed in **Table 1.1**.
- **mz error:** The error between experimental m/z value and m/z value in the library. The unit is ppm.
- **rt error:** The error between experimental rt value and rt value in the library. The unit is second (s).
- **rt score:** The match score of RT match using a trapezoidal function (**Section 1.4**). The score ranges from 0 to 1, referring to from no match to a perfect match.
- **ccs error:** The error between experimental CCS value and the CCS value in the library. The unit is percentage (%).
- **ccs score:** The match score of CCS match using a trapezoidal function (**Section 1.4**). The score ranges from 0 to 1, referring to from no match to a perfect match.
- **msms score:** The match score of MS/MS similarity between experimental MS/MS spectrum and the predicted MS/MS spectrum. It is calculated by the reverse dot-product function (**Section 1.4**). The score ranges from 0 to 1, referring to from no match to a perfect match.
- **Composite score:** The composite score is calculated by a linear combination of scores (i.e. rt score, CCS score and msms score) (**Section 1.4**). The score ranges from 0 to 1, representing the confidence level.

References:

(1) Sud, M.; Fahy, E.; Subramaniam, S. *J. Cheminf.* **2012**, 4, 23.
(2) Kind, T.; Liu, K. H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. methods* **2013**, 10, 755-758.
(3) Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z. J. *Anal. Chem.* **2017**, 89, 9559-9566.
(4) Tu J.; Yin Y.; Xu M.; Wang R.; Zhu, Z. J. *Metabolomics* **2018**, 14:5
(5) Stanstrup, J.; Neumann, S.; Vrhovsek, U. *Anal. Chem.* **2015**, 87, 9421-9428.
(6) Hines, K. M.; May, J. C.; McLean, J. A.; Xu, L. *Ana. Chem.* **2016**, 88, 7329-7336.
(7) Hartler, J.; Triebl, A.; Ziegl, A.; Trotzmuller, M.; Rechberger, G. N.; Zeleznik, O. A.; Zierler, K. A.; Torta, F.; Cazenave-Gassiot, A.; Wenk, M. R.; Fauland, A.; Wheelock, C. E.; Armando, A. M.; Quehenberger, O.; Zhang, Q.; Wakelam, M. J. O.; Haemmerle, G.; Spener, F.; Kofeler, H. C.; Thallinger, G. G. *Nat. Methods* **2017**, 14, 1171-1174.