Widely-Targeted Volatilomics (WTV) 2.0

1 General Introduction

WTV 2.0 is an open-source software with a user-friendly graphical interface, which provides a one-stop solution for the entire process of gas chromatography-mass spectrometry (GC-MS) based widely targeted volatilomics. WTV 2.0 including three modules: the *library builder* manages the library integration and deduplication, the *method generator* selects characteristic qualitative ions and generates Selective Ion Monitoring (SIM) mode acquisition method, the *data analyzer* performs the qualitative and semi-quantitative analysis of widely-targeted data.

1.1 Library builder

Library builder focusing on integrating mass spectral libraries and retention time information. It identifies and removes duplicates based on compound names, synonyms, and CAS numbers. Additionally, it performs retention time (RT) correction based on imported retention index (RI) calibration data. In addition, unknown signals in subsequent analysis can be incorporated into the library, and through similarity calculations, it prevents the import of redundant signals. The *library builder* provides the foundation for subsequent acquisition method development and data analysis.

1.2 Method generator

Method generator is responsible for development of widely targeted acquisition methods. The software selects minimum characteristic qualitative ions for each compound in the library, and allocates the selected ions into SIM segments to develop the comprehensive SIM (cSIM) acquisition method with high sensitivity and high coverage. The method generator can be also applied in the development of traditional targeted or widely-targeted acquisition methods with low coverage.

1.3 Data analyzer

Data analyzer is tailored for qualitative and semi-quantitative analysis of cSIM data. It offers a comprehensive workflow that encompasses smoothing, peak detection, peak adjustment (baseline correction and apex point correction, etc.), deconvolution, component perception, and qualitative and semi-quantitative analysis. The software provides a variety of parameter options to accommodate diverse requirements for data analysis. It also features a user-friendly interface, facilitating users to inspect analysis results and conduct data mining.

2 Using the Software

2.1 Building a Compound Library (MSP) for Detection and Identification（LibraryBuilder.exe）

2.1.1 Data Preparation

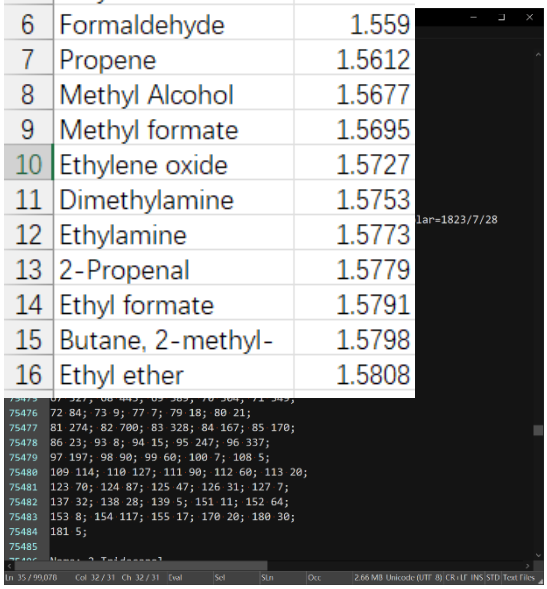
MSP file(s)

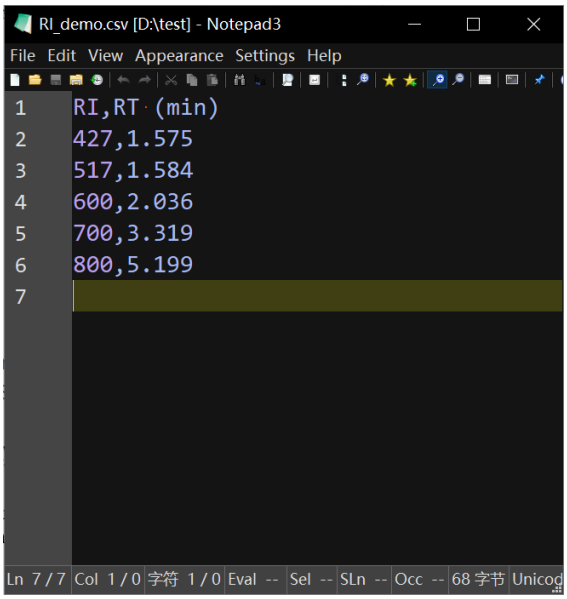
Prepare MSP files that contains mass spectrometry information and theoretical retention times for substances.

Retention Time Information

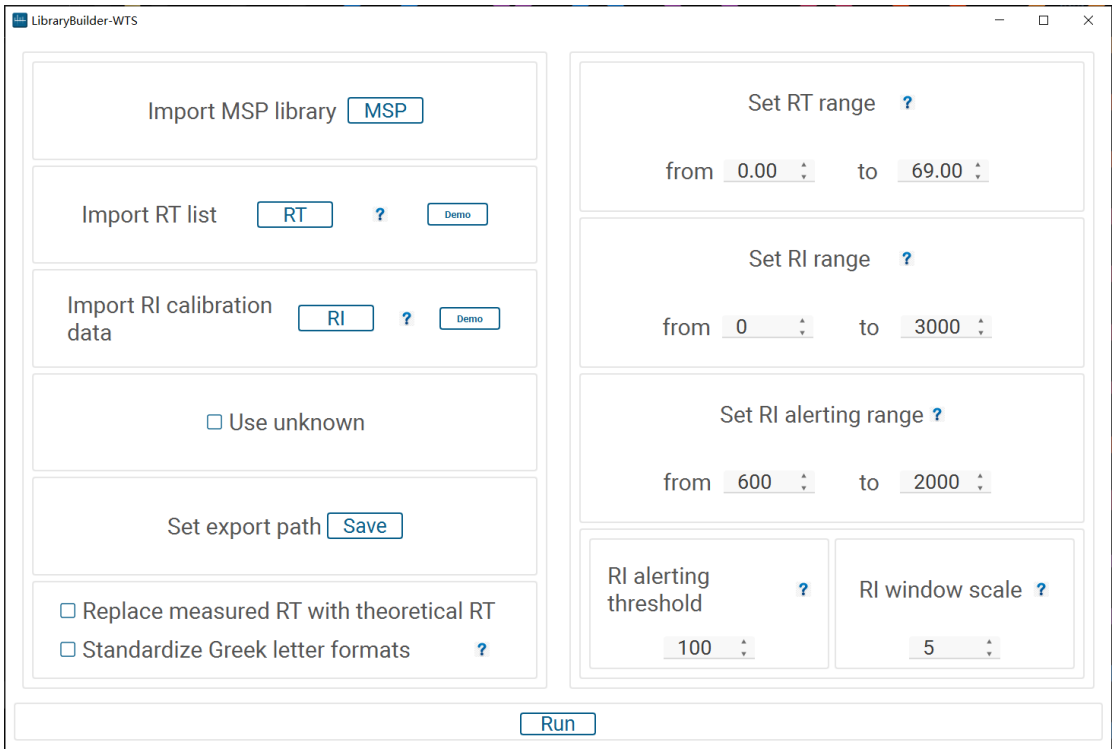
Using actual measured retention times (RT) Using RI calibration

	A	B
1	Name	RT
2	Acetaldehyde	1.496
3	Methanethiol	1.526
4	Ethane	1.5523
5	Ethyl Chloride	1.555
6	Formaldehyde	1.559
7	Propene	1.5612
8	Methyl Alcohol	1.5677
9	Methyl formate	1.5695
10	Ethylene oxide	1.5727
11	Dimethylamine	1.5753
12	Ethylamine	1.5773
13	2-Propenal	1.5779
14	Ethyl formate	1.5791
15	Butane, 2-methyl-	1.5798
16	Ethyl ether	1.5808





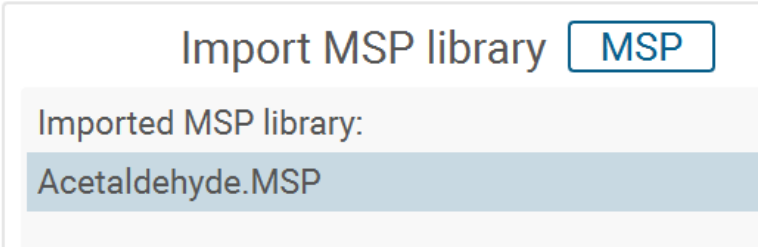
2.1.2 Parameter Configuration



Import MSP library

To input an MSP file, simply click the 'MSP' button and select the desired MSP file.

Once selected, the interface will display all the selected file directories, confirming a successful input.

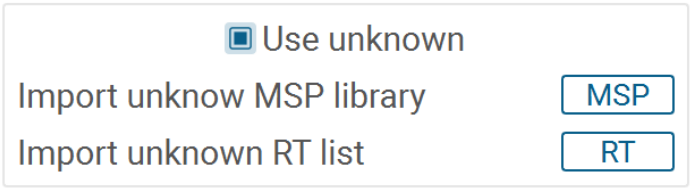


Import RT list: import a list of compounds and their RT.

Import RI calibration data: import RI calibration data.

Use unknown: You can use this feature when you need to enhance signals from compounds that cannot be qualitatively identified.

When you wish to use this feature, simply select the white box to highlight it, and the hidden menu will become visible.



Within the hidden menu, input the corresponding information for the unknown item.

Set export path

Click the 'Save' button, choose a saving path, and the processed results will be saved to this directory.

Replace measured RT with theoretical RT

Selecting this option will replace measured retention times outside the set theoretical retention time window with the theoretical retention time.

Standardize Greek letter formats

This option allows for a unified representation of Greek letters. Eg: ".alpha." -> "alpha"

Set RT range

Once configured, the output retention times will be constrained within this range.

Set RI range

Once configured, the output retention index will be constrained within this range.

Set RI alerting range

Only unreasonable RT (Retention Time) values outside the configured range will trigger a warning.

RI alerting threshold

Compound will be marked if the difference between the measured RI and the database Rl exceeds this threshold.

RI window scale

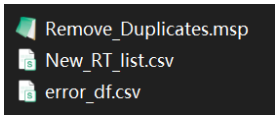
The larger this value, the larger the Rl error threshold when the RI is larger. Setting it to 0 disables this feature. For more specific and detailed information, please refer to the official documentation of the AMDIS software.

Tips:

Any file directory entered can be deleted by double-clicking the entered entry.

2.1.3 Results and Reports

After configuring all parameters, click the 'Run' button. The program will automatically generate three result files.



The MSP result file:

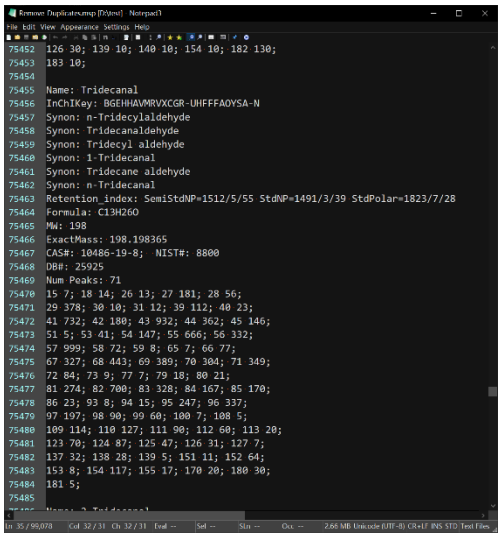
The RT (Retention Time) result file

	A	B	C	D	E
1	Name	RT	RI_msp	RI_input	Alert
842	5-Nonanol	21.528	1092	1095	
843	2-Furanmethanol, propar	21.5884	1096	The content of the retention time actually detected was not retrieved	rt_is_in_silico
844	Fenchone	21.5884	1096	The content of the retention time actually detected was not retrieved	rt_is_in_silico
845	Pyrazine, 2-methoxy-3-(1	21.6555	1097	The content of the retention time actually detected was not retrieved	rt_is_in_silico
846	2-Undecene, (E)-	21.6555	1097	The content of the retention time actually detected was not retrieved	rt_is_in_silico
847	7-Octen-4-ol, 2-methyl-	21.6555	1097	The content of the retention time actually detected was not retrieved	rt_is_in_silico
848	o-Guaiacol	21.689		1097	
849	Guaicol	21.7		1098	
850	Pentanoic acid, butyl este	21.7227	1098	The content of the retention time actually detected was not retrieved	rt_is_in_silico
851	3-(Methylthio)propanoic	21.7227	1098	The content of the retention time actually detected was not retrieved	rt_is_in_silico
852	trans-Linalool oxide (furar	21.743	1086	1098	
853	3-Nonanol	21.754	1095	1098	
854	4-Nonanol	21.808	1088	1099	
855	2H-Pyran-2-one, tetrahyc	21.817	1092	1099	
856	Methyl benzoate	21.87		1100	
857	Benzoic acid, methyl ester	21.915	1094	1101	
858	Butanoic acid, 2-methyl-,	21.9263	1101	The content of the retention time actually detected was not retrieved	rt_is_in_silico
859	6-Nonenal, (Z)-	21.9263	1101	The content of the retention time actually detected was not retrieved	rt_is_in_silico
860	2-Cyclohexen-1-one, 4,4	21.9263	1101	The content of the retention time actually detected was not retrieved	rt_is_in_silico
861	Furan, 3-(4-methyl-3-pei	21.9263	1101	The content of the retention time actually detected was not retrieved	rt_is_in_silico
862	2-Nonanone	21.94	1092	1101	
863	Caryophyllenyl alcohol	21.95		1101	
864	3-Hexen-1-ol, propanoat	21.962	1100	The content of the retention time actually detected was not retrieved	rt_is_in_silico
865	(Z,Z)-3,6-Nonadienal	21.962	1100	The content of the retention time actually detected was not retrieved	rt_is_in_silico
866	2-Undecene, (Z)-	21.9955	1102	The content of the retention time actually detected was not retrieved	rt_is_in_silico
867	Diallyl disulphide	21.999	1081	1102	
868	Thujone	22.0648	1103	The content of the retention time actually detected was not retrieved	rt_is_in_silico
869	(E)-1-Allyl-2-(prop-1-en	22.0648	1103	The content of the retention time actually detected was not retrieved	rt_is_in_silico
870	1-Nonen-4-ol	22.0648	1103	The content of the retention time actually detected was not retrieved	rt_is_in_silico
871	Decane, 2,6,8-trimethyl-	22.134	1104	The content of the retention time actually detected was not retrieved	rt_is_in_silico
872	2,2,6-Trimethyl-3-keto-6	22.186		1105	
873	Butanoic acid, 2-methyl-,	22.2033	1105	The content of the retention time actually detected was not retrieved	rt_is_in_silico
874	Ethanone, 1-(4,5-dihydro	22.2726	1106	The content of the retention time actually detected was not retrieved	rt_is_in_silico
875	3-Cyclohexene-1-methar	22.2726	1106	The content of the retention time actually detected was not retrieved	rt_is_in_silico
876	Disulfide, dipropyl	22.3418	1107	The content of the retention time actually detected was not retrieved	rt_is_in_silico
877	1,3,8-p-Menthatriene	22.353	1119	1107	
878	Unknown S	22.376		1107	
879	Propanoic acid, hexyl este	22.4111	1108	The content of the retention time actually detected was not retrieved	rt_is_in_silico
880	2H-Pyran-3(4H)-one, 6-e	22.4111	1108	The content of the retention time actually detected was not retrieved	rt_is_in_silico
881	2,6,6-Trimethylbicyclo[3.2	22.4111	1108	The content of the retention time actually detected was not retrieved	rt_is_in_silico
882	Phenol, 2,6-dimethyl-	22.4111	1108	The content of the retention time actually detected was not retrieved	rt_is_in_silico
883	Pentanoic acid, 3-methylb	22.4111	1108	The content of the retention time actually detected was not retrieved	rt_is_in_silico
884	Filifolone	22.42		1108	
885	Hexanoic acid, propyl este	22.445	1094	1108	

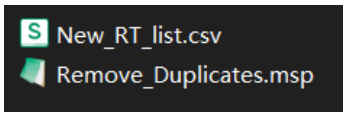
The warning information file

	A	B	C	D	E	F	G	H	I	J
1	Name	reason								
2	n-Hexadecanoic acid	WARNING: This compound already exists in the database. Duplicates are not allowed.								
3	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
4	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
5	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
6	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
7	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
8	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
9	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								
10	Phenylethyl Alcohol	WARNING: This compound already exists in the database. Duplicates are not allowed.								

Warning information including:



WARNING: The compound was not found in the provided MSP file.
WARNING: This compound already exists in the database. Duplicates are not allowed.
WARNING: The RT value is out of the valid range.
WARNING: The synonym name has been changed to NIST Preferred Name.



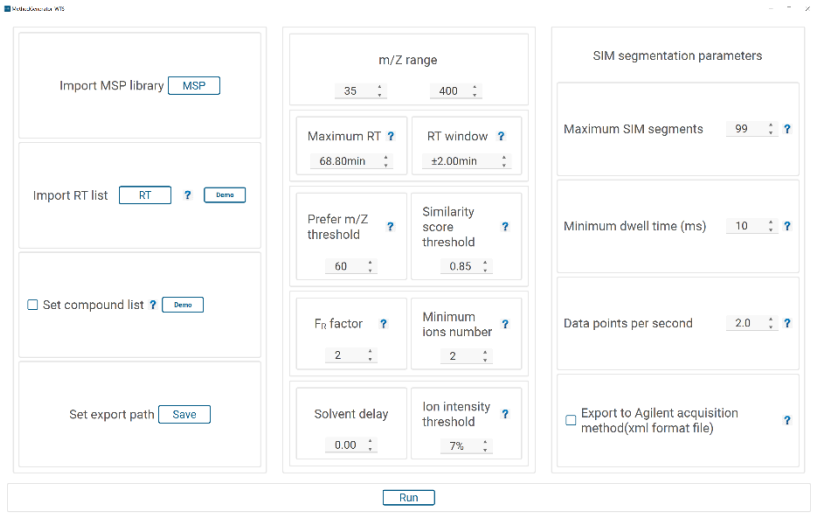
	A	B
1	Name	RT
2	Acetaldehyde	1.496
3	Methanethiol	1.526
4	Ethane	1.5523
5	Ethyl Chloride	1.555
6	Formaldehyde	1.559
7	Propene	1.5612
8	Methyl Alcohol	1.5677
9	Methyl formate	1.5695
10	Ethylene oxide	1.5727
11	Dimethylamine	1.5753
12	Ethylamine	1.5773

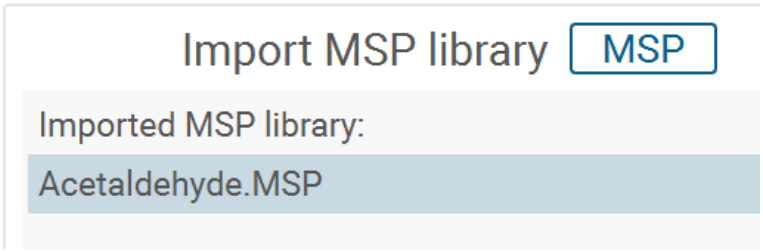
2.2 Generating a Widely-Targeted SIM Method for Compound Detection（MethodGenerator.exe）

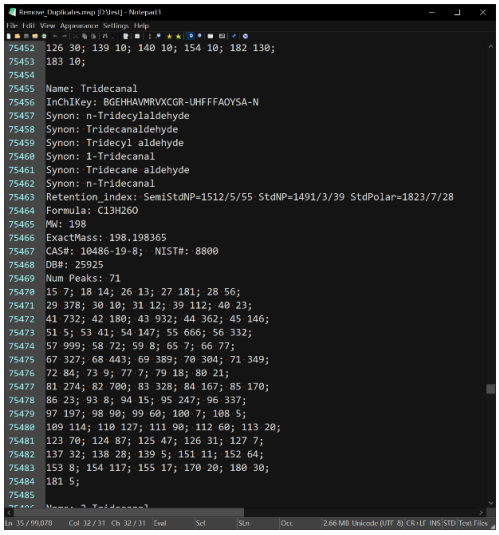
2.2.1 Data Preparation

Before using this software, it is necessary to prepare two input files. The input files can directly utilize the corresponding result files generated by LibraryBuilder.
These input files should include the compound names and their corresponding RT (Retention Time) information.

In addition, you should also have a Mass Spectrometry (MS) information file in the MSP format.







2.2.2 Parameter Configuration

Import MSP library

To input an MSP file, simply click the 'MSP' button and select the desired MSP file.

Once selected, the interface will display all the selected file directories, confirming a successful input.

Import RT list: Import a list of compounds and their RT.

Set compound list: This feature is used when you want to detect only a subset of compounds included in the MSP file, rather than all of them. If you choose this option, the generated target will only include the compounds mentioned in the TXT file.

Set export path:

Click the 'Save' button, choose a saving path, and the processed results will be saved to this directory.

m/z range:

The qualitative ions will be selected from within this range.

Maximum RT: Enter the end time of the temperature ramp program.

RT window: When selecting qualitative ions, target compounds are compared for similarity with substances within the user-defined RT (Retention Time) window.

Prefer m/z threshold: M/Z values below this threshold will be assigned a lower weight in the calculation of the weighted score. The default threshold is set to 60.

Similarity score threshold: When the similarity score is below the threshold of 0.85, two spectra are considered distinguishable. The default threshold is set to 0.85.

F_R factor : The 'Ratio of Peak Pairs' term is considered in the similarity score calculation only if the number of ions is greater than the specified threshold. The F_R factor is typically set to be consistent with the minimum number of ions. For more specific and detailed information, please refer to the official documentation of the AMDIS software.

Minimum ions number: This represents the minimum number of ions required for the selection of qualitative ions.

Solvent delay: Enter the start time for mass spectrometry detection, also known as the solvent delay time.

Ion intensity threshold: Ions with abundances lower than the maximum ion abundance multiplied by this threshold value will be excluded from the selection of qualitative ions.

SIM segmentation parameters:

Maximum SIM segments: The maximum number of segments allowed in the SIM (Selected Ion Monitoring) acquisition method. Default: 99.

Minimum dwell time (ms): The allowed minimum dwell time. Default: 10 ms.

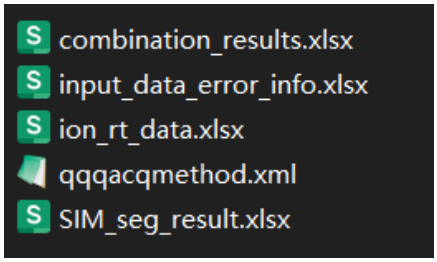
Data points per second: Adjust the number of data points per second. If the calculated dwell time based on the specified data points per second falls below the minimum dwell time, the data points per second will be decreased to ensure that the dwell time remains greater than or equal to the minimum dwell time. The default setting is 2.

Export to Agilent data acquisition method (xml format file): Select this option to export the XML-formatted file used by Agilent data acquisition software.

```xml
<?xml version='1.0' encoding='UTF-8'?>
<MSAcqMethod xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xm
  <msInstrument>QQQ</msInstrument>
  <ionSource>EI</ionSource>
  <tuneFile>atunes.eiex.tune.xml</tuneFile>
  <stopMode>ByChromatographTime</stopMode>
  <stopTime>1</stopTime>
  <solventDelay>0</solventDelay>
  <collisionGasOn>true</collisionGasOn>
  <sourceParameters>
    <sourceParameter>
      <id>SourceHeater</id>
      <posPolarityValue>250</posPolarityValue>
      <negPolarityValue>250</negPolarityValue>
    </sourceParameter>
  </sourceParameters>
  <isTimeFilterEnabled>true</isTimeFilterEnabled>
  <timeFilterPeakWidth>0.0133333337</timeFilterPeakWidth>
  <timeFilter>
    <activeCount>1</activeCount>
    <definition>
      <time>0</time>
      <peakWidth>0.0133333337</peakWidth>
    </definition>
    <definition>
      <time>10</time>
      <peakWidth>0.05</peakWidth>
    </definition>
  </timeFilter>
```

2.2.3 Generated Widely-Targeted SIM Method

combination_results: This displays the combined information of target compounds after merging.



input_data_error_info: This file highlights discrepancies in the input data that led to improper identification and analysis.

ion_rt_data: This file displays the results of the final compounds and their qualitative ions.

The content of this file is identical to the 'combination_results' file, but it is in a different format to

	RT	Ion_Combination	Note	Similar_Compound_List	SCL_Note
Acetaldehyde	1.496	[43, 44]		['Methyl formate']	
Methanethiol	1.526	[47, 48]		[]	

make it more user-friendly for other purposes.

SIM_seg_result: This file presents the results of the SIM segments.

qqqacqmethod.xml: The XML-formatted file used by Agilent data acquisition software.

	RT	ion
Acetaldehyde	1.496	43
Acetaldehyde	1.496	44
Methanethiol	1.526	47
Methanethiol	1.526	48
Ethyl Chloride	1.555	64
Ethyl Chloride	1.555	66
Propene	1.5612	41
Propene	1.5612	40
Propene	1.5612	37
Propene	1.5612	39

2.3 Data Analysis

2.3.1 Data Preparation

2.3.1.1 Mass Spectrometry File Format Conversion

Before data analysis, we need converse raw data to open formats (.mzML/.cdf) by open access application.

Data conversion to open format (.mzML)

Conversion with MSConvert

Download and install ProteoWizard from here (http://proteowizard.sourceforge.net/downloads.shtml) .

After installation, from the Start Menu, click the ProteoWizard folder and open MSConvert.

Click Browse and select file(s) for conversion. Then click Add to add them to the MSConvert workflow.

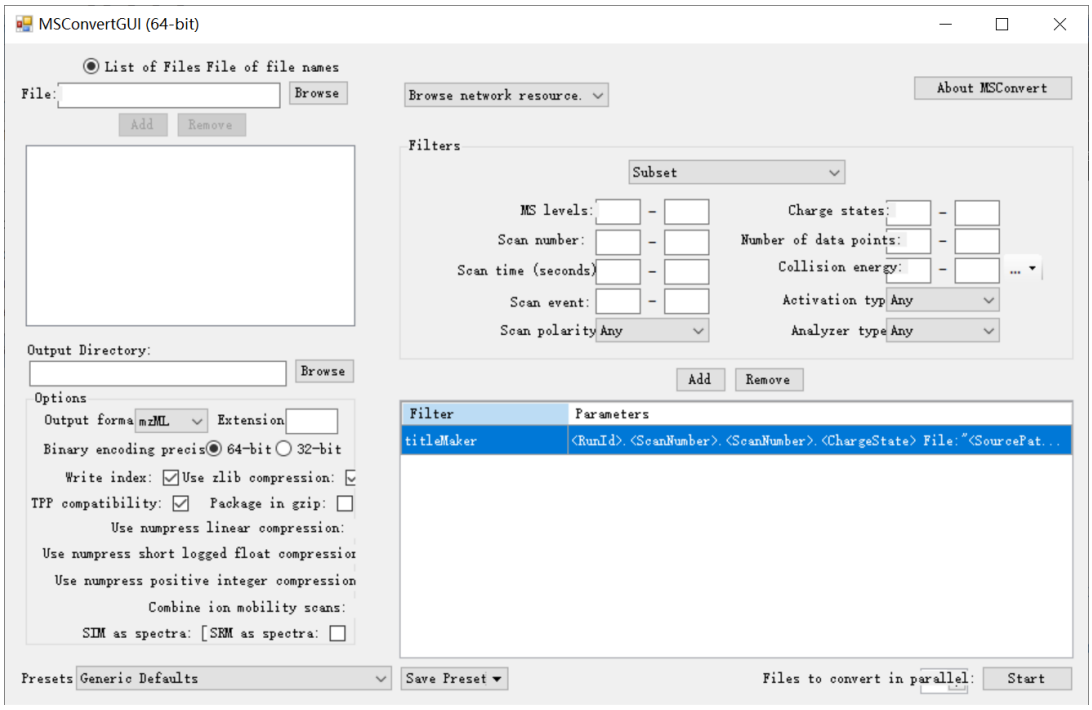
Choose an Output Directory

Under Options, choose mzML (prefered) or mzXML for output format, 32-bit for binary encoding precision, SIM as spectra and uncheck Use zlib compression.

Click Add to add the filter.

Save the parameters for the next conversion. This will save you some time and prevent misconfiguration. In Presets (left bottom), click on Save Presets, and select "Save as default for the format".

Click on Start. Check your folder for the new .mzML files. Verify that these files open properly in Insilicos or TOPP View (OpenMS http://www.openms.de/).

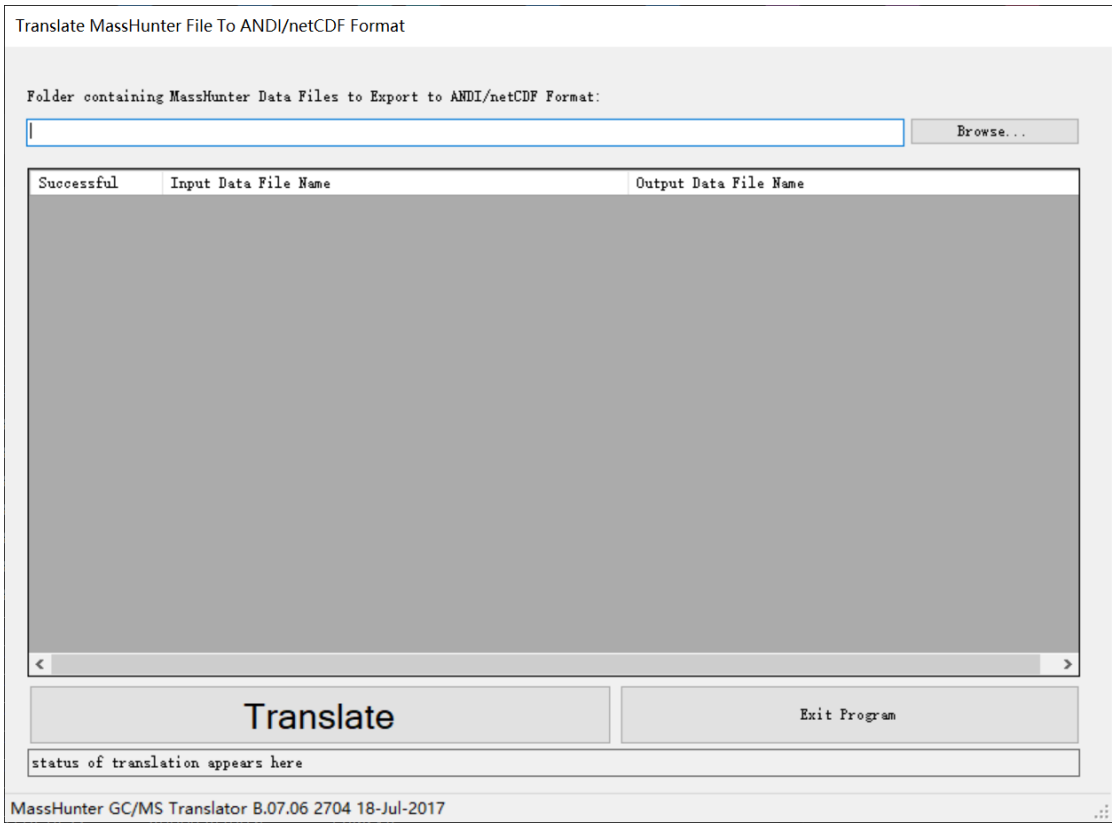


Data conversion to open format (.cdf)

Agilent: Click Browse and select folder containing file(s) for conversion.

Click Translate

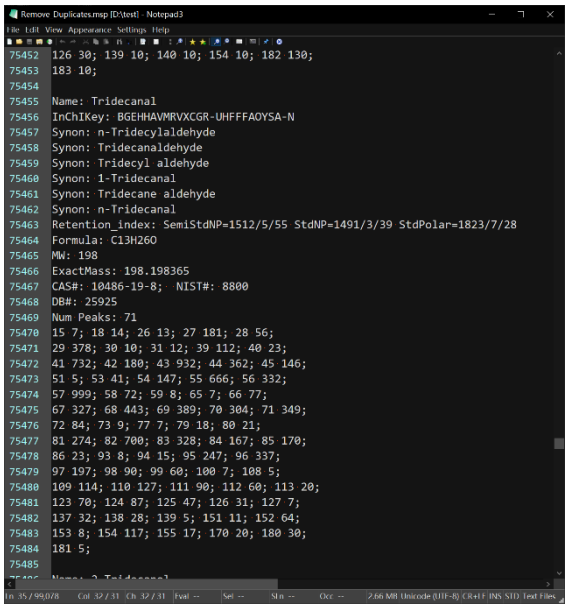
The CDF files are generated in the same folder as the raw data files.



For data processing methods from other vendors, please refer to the relevant data conversion instructions.

2.3.1.2 Compound Library (MSP) for identification

Input an MSP file that contains all the target metabolites. The input files can directly utilize the



corresponding result files generated by LibraryBuilder.

2.3.1.3 Optional: Retention Time Information

To identify compounds using actual measured retention times (RT), you will need the required files. The input files can directly utilize the corresponding result files generated by LibraryBuilder.

	A	B
1	Name	RT
2	Acetaldehyde	1.496
3	Methanethiol	1.526
4	Ethane	1.5523
5	Ethyl Chloride	1.555
6	Formaldehyde	1.559
7	Propene	1.5612
8	Methyl Alcohol	1.5677
9	Methyl formate	1.5695
10	Ethylene oxide	1.5727
11	Dimethylamine	1.5753
12	Ethylamine	1.5773
13	2-Propenal	1.5779
14	Ethyl formate	1.5791
15	Butane, 2-methyl-	1.5798
16	Ethyl ether	1.5808

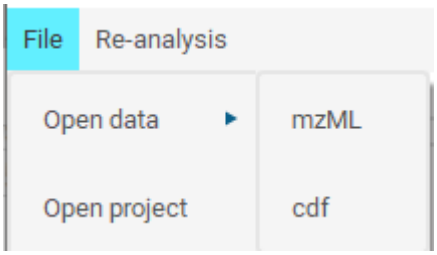
To identify compounds using Kovats Index calculation, you will need the necessary files.

	A	B
1	Num	RT (min)
2	8	5.219
3	9	9.34
4	10	15.177
5	11	21.877
6	12	28.764
7	13	35.514
8	14	41.988
9	15	48.144
10	16	53.937
11	17	59.401
12	18	63.787
13	19	64.851
14	20	65.484

2.3.2 Parameter Configuration

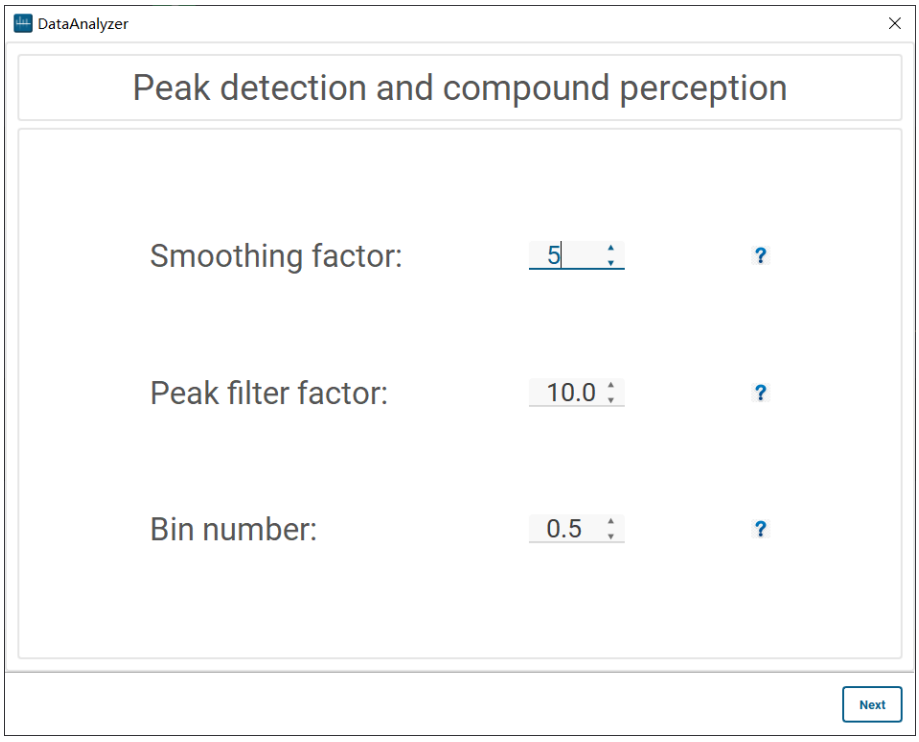
2.3.2.1 Data Import

Select the appropriate option based on the transformed file format.



2.3.2.2 Peak Detection

Smoothing factor:
A high smooth factor can make the peaks smoother, but it may also lead to the loss of low-abundance peaks. The default value is set to 5.
Peak filter factor:
A high peak filter factor can remove low-abundance peaks. If this value exceeds 10, it will significantly extend the program's runtime. The default setting is 10.
Bin number:
When using a low data point acquisition mode, such as two points per second, it is recommended to set a wide peak grouping width, for example, 0.5. For more specific and detailed information, please refer to the official documentation of the AMDIS software.



2.3.2.3 Qualitative Analysis

In RI mode

Library search window: Configure the retrieval target compound window in RI mode. The program will compare all compounds within the specified matching window.

Maximum RI: The maximum RI (Retention Index) value.

Peak group Match weight: Peak group forward retrieval matching score weight.

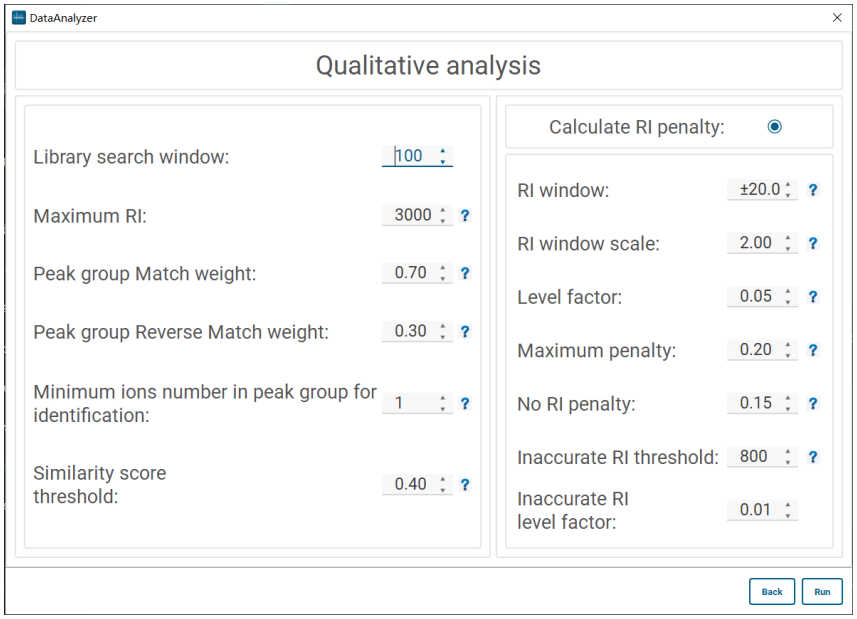
Peak group Reverse Match weight: Peak group reverse retrieval matching score weight

Minimum ions number in peak group for identification:

The minimum number of ions required for a peak group.

Similarity score threshold: The integrated similarity score threshold, where only compounds with scores above this threshold will be considered as candidate compounds.

Calculate RI penalty: Selecting this option will penalize candidate compounds with a significant



difference in RI values.

RI window: Set the RI window range in which no penalty will be applied.

RI window scale: The RI penalty window will be scaled linearly by this factor. Setting it to 0 disables this feature. The default value is 2.

Level factor: A higher number indicates a more severe penalty.

Maximum penalty: Set the maximum penalty score.

Inaccurate RI threshold: RI values below this threshold will be affected by the Inaccurate RI level factor.

Inaccurate RI level factor: The specified range for the Inaccurate RI threshold mentioned above is used for the following purposes.

In RT mode

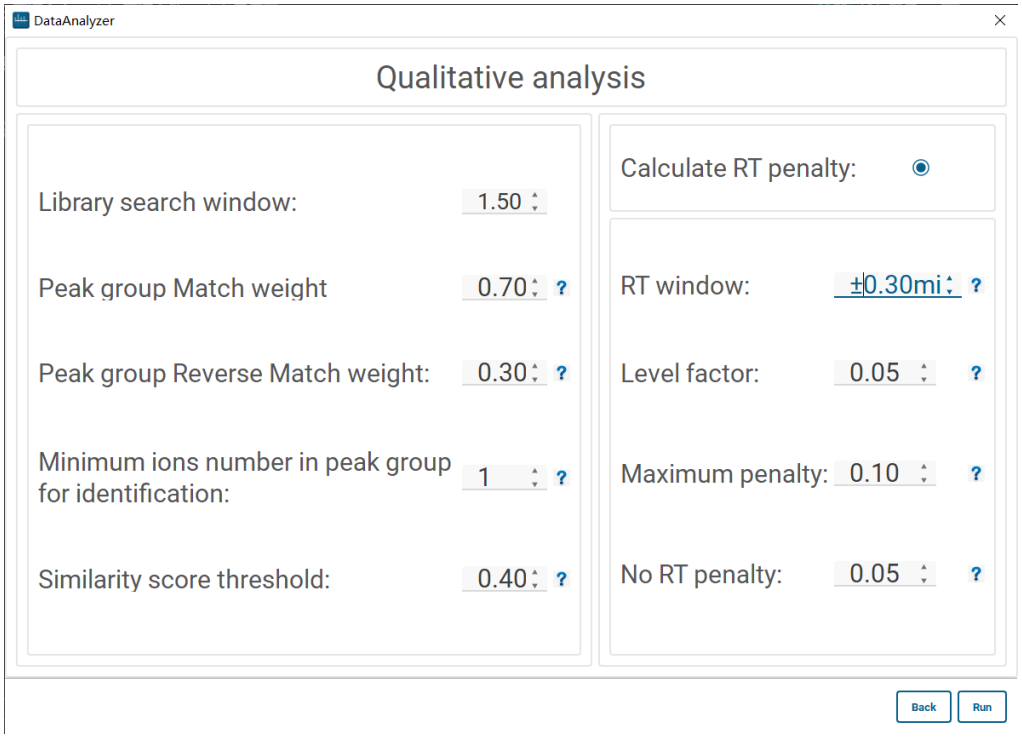
Library search window: Configure the retrieval target compound window in RT mode. The program will compare all compounds within the specified matching window.

Peak group Match weight: Peak group forward retrieval matching score weight.

Peak group Reverse Match weight: Peak group reverse retrieval matching score weight.

Minimum ions number in peak group for identification:

The minimum number of ions required for a peak group.



Similarity score threshold: The integrated similarity score threshold, where only compounds with scores above this threshold will be considered as candidate compounds.

Calculate RT penalty: Selecting this option will penalize candidate compounds with a significant difference in RT values.

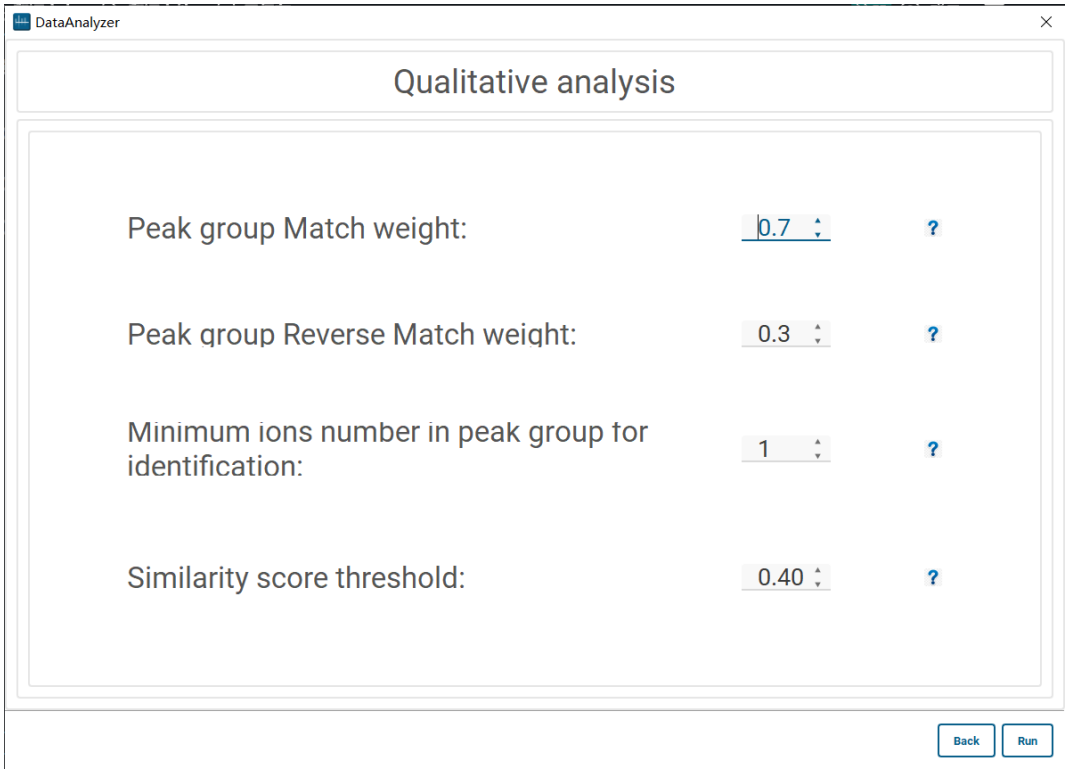
RT window: Set the RT window range in which no penalty will be applied.

Level factor: A higher number indicates a more severe penalty.

Maximum penalty: Set the maximum penalty score.

No RT penalty: Peak groups with retention times (RT) lower than this value will not receive a penalty for mismatch.

Choose the mode for not entering retention time information.



Peak group Match weight: Peak group forward retrieval matching score weight.

Peak group Reverse Match weight: Peak group reverse retrieval matching score weight.

Minimum ions number in peak group for identification:

The minimum number of ions required for a peak group.

Similarity score threshold: The integrated similarity score threshold, where only compounds with scores above this threshold will be considered as candidate compounds.

2.3.3 Main Workspace Overview