MS-DIAL software parameters for processing untargeted metabolomics data acquired on the ZenoTOF 7600 system

Cagakan Ozbalci[1], Paul RS Baker[2], and Rebekah Sayers[1]

[1]SCIEX, UK; [2]SCIEX, USA

This technical note demonstrates the importance of parameter settings in MS-DIAL version 4.92 to process untargeted metabolomic data acquired using the ZenoTOF 7600 system. Previous technical notes[1-2] have shown the importance of instrumental settings for data acquisition and best practices for data processing using SCIEX OS software to produce the highest quality data with broad metabolome coverage. However, different software programs have different algorithms that process data with uniquely defined parameters and settings (**Figure 1)**. Furthermore, parameter settings can be instrument-dependent, which requires comprehensive testing to determine the best parameter settings for a particular data set. Herein, MS-DIAL 4.92 software[3], a widely used processing tool for metabolomics and lipidomics data analysis, was used to interpret untargeted metabolomics data. Parameter settings were sequentially adjusted through iterative data processing to reveal how setting changes affect metabolomics results. Optimal MS-DIAL software parameter setting values are presented for data acquired using the ZenoTOF 7600 system.

Key features of optimizing MS-DIAL software parameters for processing metabolomics data acquired using the ZenoTof 7600 system

- MS-DIAL software rapidly processes untargeted metabolomics data acquired on the ZenoTOF 7600 system
- The high sensitivity of the ZenoTOF 7600 system enables low threshold settings within MS-DIAL software that improves the overall coverage of the metabolomics experiment
- MS-DIAL software is compatible with most metabolomic compound libraries

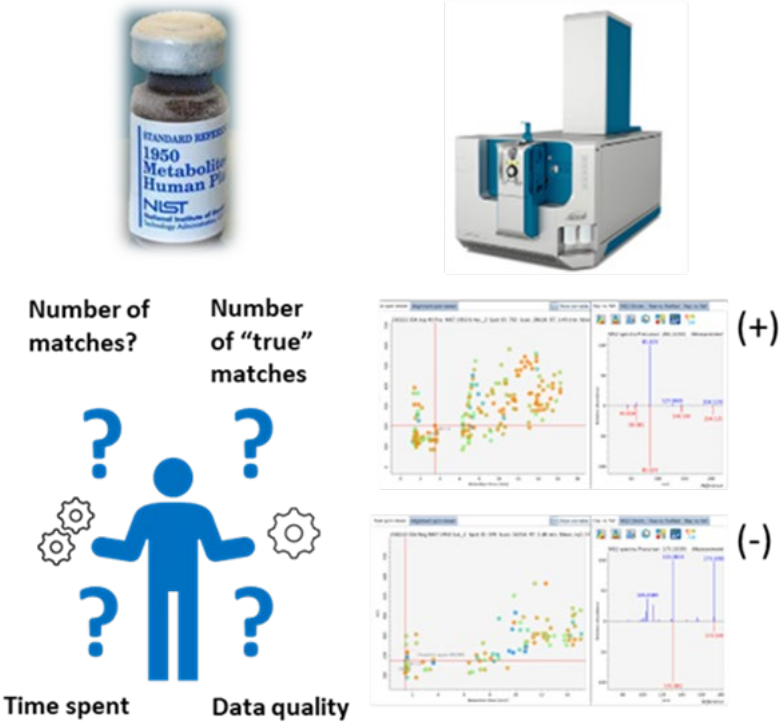

Figure 1. A major bottleneck for untargeted metabolomics workflows is time spent data processing. Understanding the software-specific parameter settings and how they should be adjusted to accommodate specific instrument data is essential to maximize coverage while maintaining high confidence in a timely manner. Herein optimal parameter settings for MS-DIAL 4.92 software are described for untargeted metabolomics data acquired on the ZenoTOF 7600 system.

Introduction

Untargeted metabolomics aims to detect and quantify all observable small biomolecules within a sample to define the metabolic state of an organism and potentially identify biomarkers of disease[4]. High-resolution mass spectrometry (HRMS) analysis using a data-dependent acquisition (DDA) mode is the primary tool for untargeted metabolomics experiments. To increase compound detection and identification (i.e., coverage), experiments are typically run in both positive and negative ion modes. Acquired data are generally processed using software that matches MS/MS spectra to small molecule databases for compound identification.

The breadth of coverage of metabolic compounds by mass spectrometry depends on several factors. First, instrument performance and hardware parameter settings affect metabolomics data quality. Optimal parameter settings for the ZenoTOF 7600 system have been previously reported[1] that leverage the speed and sensitivity of the instrument to maximize coverage from a DDA experiment. Second, data processing parameters significantly impact the identification of metabolites. These parameters can be related to library match score, minimum intensity threshold, etc. Incorrect parameter settings may result in misidentification or missing a compound altogether. Software parameter settings for SCIEX OS software for untargeted metabolomics data acquired using the ZenoTOF 7600 system have been previously determined[2]. MS DIAL software can also process these data; however, its parameter settings are unique from SCIEX OS software. Systematic adjustment of parameter settings and data review are required to find the optimal software parameter settings.

In this technical note, untargeted metabolomics data were acquired from NIST SRM plasma samples using the ZenoTOF 7600 system. Data were processed by MS DIAL 4.92 software using iterative permutations of different parameter settings for *minimum threshold, mass slice width,* and *identification cut-off* tabs in the processing workflow. From these processed data, optimal software processing parameters were identified and are reported herein.

Methods

Sample preparation: NIST SRM 1950 samples were extracted by a one-phase liquid extraction. Four volumes of ice-cold ethanol were added to 1 sample volume and vortexed for 30 seconds. Extraction mixtures were centrifuged to separate the precipitated protein debris, and the supernatant was used directly for metabolomics analysis. The supernatant can be stored at -20 °C for future analysis.

Chromatography: Extracted metabolites were resolved using an Exion UHPLC instrument equipped with a Kinetex F5 column (2.1 × 150 mm, 2.6 µm; Phenomenex). The column oven temperature was 40°C with a constant flow rate of 0.2 mL/min. Gradient details are shown in **Table 1.**

Table 1: Chromatographic gradient (flow rate = 0.20 mL/min)

Time (min)	Mobile phase A (%)	Mobile Phase B (%)
0.0	100	0
2.1	100	0
14	5	95
16	5	95
16.1	0	0
20	0	0

Mass spectrometry: Extracted samples were analyzed using a ZenoTOF 7600 system with an OptiFlow Turbo V ion source. A top-40 DDA method with dynamic background subtraction (DBS) and exclusion for 6 s after 3 occurrences was employed for both positive and negative ion modes. Electrospray ionization voltages were set to 5500 V and -4500 V for positive and negative ion modes, respectively. The collision energy (CE) was set to 30 V for positive and -25 V for negative ion mode, and the accumulation time was set to 5ms. The automated calibrant delivery system (CDS) performed automated calibration every 9 samples. A summary of the MS instrument parameters is presented in **Table 2**.

Table 2: ZenoTOF 7600 system source and gas parameters

Parameter	Setting (Pos)	Setting (Neg)
Curtain gas (CUR)	35	35
Ion source gas 1 (GS1)	50	50
Ion source gas 2 (GS2)	70	70
CAD Gas (CAD)	7	7
Source temperature (TEM)	500 °C	500 °C
Ion spray voltage (IS)	5500 V	-4500 V
TOF MS mass range	60-1000 Da	60-1000 Da
Declustering Potential (DP)	40	-40
Time bins to sum	6	6
Accumulation time	5 ms	5 ms
IDA max candidate ions	40	40
Dynamic background subtraction (DBS)	Yes	Yes
Collision energy (CE)	30	-25
Zeno pulsing	Yes	Yes
Zeno threshold	20000	20000

Data processing: All data were processed using MS-DIAL 4.92 open-source software. For identification, two different libraries (*ExpBioInsilico_NEG_VS17.msp* and *ExpBioInsilico_Pos_VS17.msp*) were employed. These libraries can be downloaded from (http://prime.psc.riken.jp/compms/msdial/main.html) and can easily be selected in the "Identification" tab of the software.

Project creation and analysis of parameter settings

Project creation

Initial "new project" creation is illustrated in **Figure 2**. Selecting the project file path where the raw files are stored is critical. WIFF and WIFF2 files can be submitted to MS-DIAL without conversion. Click "Soft ionization," "Chromatography," and "Conventional LC/MS or dependent MS/MS" for the first three tabs. Other noteworthy options to consider are "Data type" (MS1 and MS2). Choose "Centroid data" for both sections. Lastly, select "Metabolomics" for the target omics and "negative" or "positive" for the ion mode.

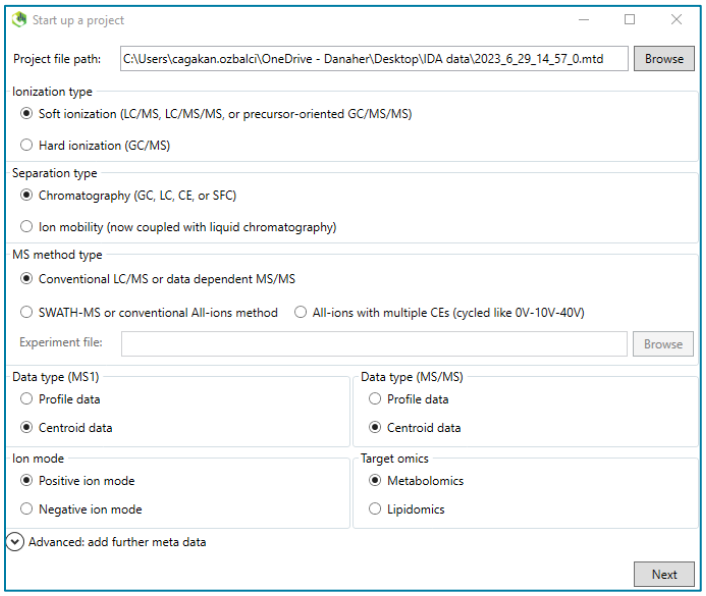

Figure 2. New project creation page

Analysis parameter settings

Data collection tab:

In this section, MS-DIAL software parameter settings critically affect coverage and the time needed to process the data files. Significantly, these settings also affect the number of false positive results. In the "Mass accuracy" settings, the default values are recommended. Next, click the "advanced" button, as shown in **Figure 3**. To limit the data processing mass range and potentially reduce the time required for processing, enter the appropriate MS1 and MS2 mass ranges. It is also recommended to define when retention time starts and ends to remove unnecessary data points, such as the washing step, from the analysis. The other settings can be kept as default values. If a powerful workstation computer is used, the "Number of threads" setting can be set to more than 2.

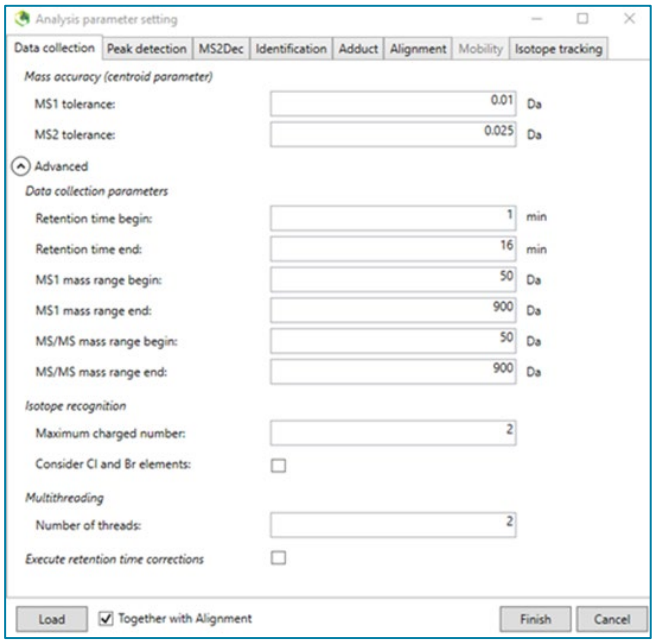

Figure 3. Data collection tab with advanced settings

Peak detection tab:

Under the "Peak detection" tab, there are two parameters to be adjusted for optimal peak detection: "Minimum peak height" and "Mass slice width" (**Figure 4**). One of the goals of this technical note was to investigate how varying these parameters affects the data quality and the overall processing time. For "minimum peak height," 50, 250, 500, and 1000 amplitude were sequentially selected, and the mass slice was set to 0.05 or 0.1 Da. If the smoothing method needs to be changed, click Advanced in this section for detailed smoothing settings.

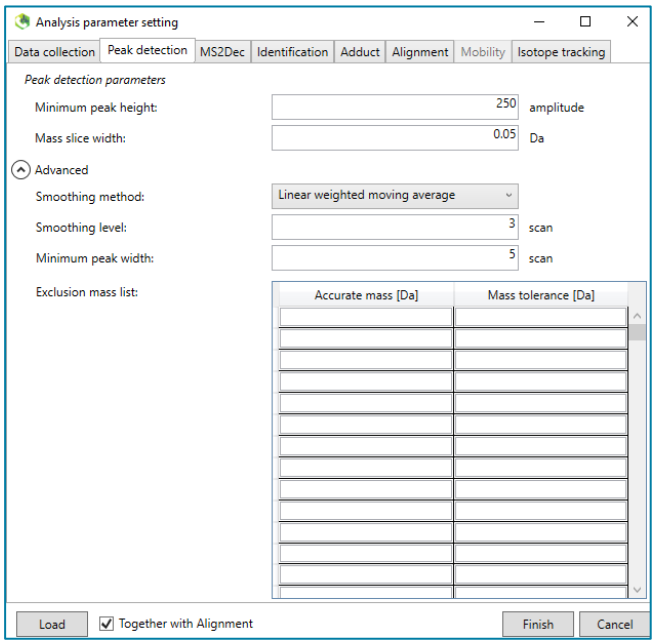

Figure 4. Peak detection settings tab with smoothing method details

Identification tab:

In the identification section, an MSP file, which can be downloaded from MS-DIAL's website, should be selected, and the identification cut-off value should be set. The default value for this parameter is 80%, but in this study, values of 70% and 60% were also used to observe the effect on the final results. The remaining parameter settings can be set as shown in Figure 5. If a custom library is to be used, click on the advanced button and select the library. (The library should be stored as a *.txt file.)

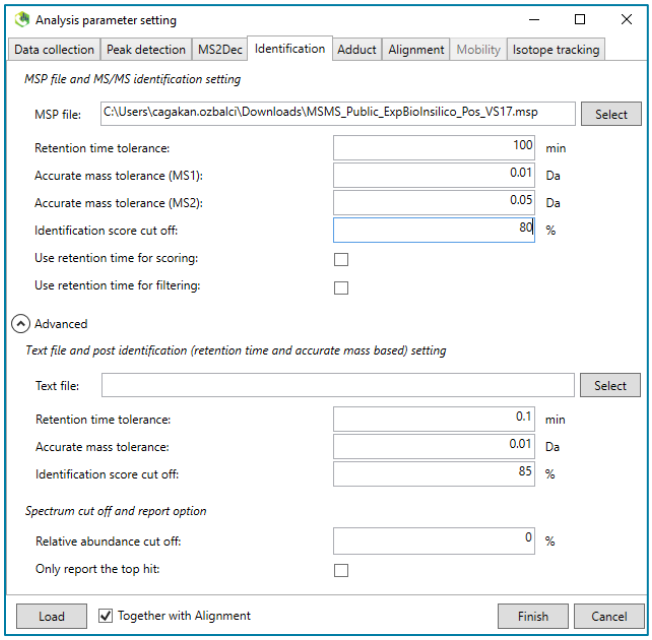

Figure 5. Identification tab

Adducts tab:

Metabolite adducts can vary depending on the modifiers added to solvents. For example, in the positive ion mode, a compound can appear as a protonated ion or as a sodium, potassium, or ammonium adduct. The recommended adducts to be selected in both positive and negative ion modes are presented in **Figure 6**.

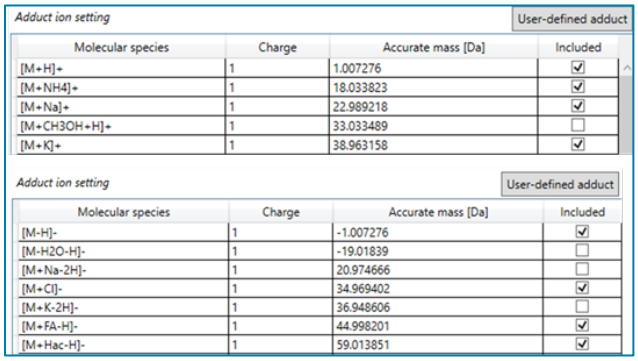

Figure 6. Adduct ion setting tab

Alignment tab:

This is the final section to be completed before processing a sample or a sample batch (**Figure 7**). A reference sample file should be chosen from a pooled QC sample or from the sample with the highest metabolite concentration. The remaining parameters (and those under the Advanced tab) were set to the default values since they are compatible with the ZenoTOF 7600 system data acquisition. Once all parameters have been set, click "Finish" and wait until the results screen appears.

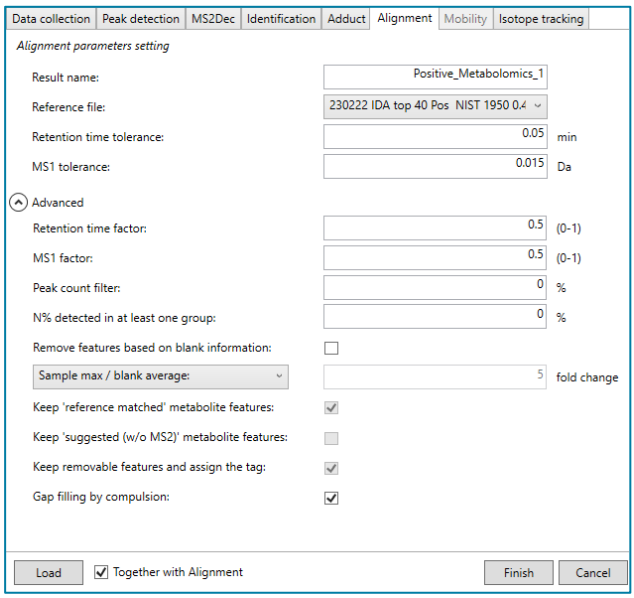

Figure 7. Alignment tab

Results and discussion

Metabolites extracted from NIST SRM 1950 plasma samples were analyzed in positive and negative ion modes, with two different injection volumes for each polarity. For the positive ion mode, 0.4 µL and 2 µL of the extract were analyzed, whereas 1 µL and 5 µL were injected in the negative ion mode. **Tables 3 and 4** show how adjusting the different parameter settings affects the numbers of identified metabolites in the positive and negative ion modes. Only the higher volume injection results are displayed in these tables. For both polarities, the higher injection volume resulted in approximately 1.4 more annotated metabolites than the lower volume injections (data not shown).

Table 3. Coverage, time for analysis, and confidence under different parameter settings in the positive ion mode.

Polarity	Volume	Min Threshold	Mass slice widh	Identification cut off	Number of hits	Time for analysis
Positive	2 µL	50	0.1	60%	573	Very long
Positive	2 µL	50	0.05	60%	651	Very long
Positive	2 µL	250	0.1	60%	505	Long
Positive	2 µL	250	0.05	60%	556	Long
Positive	2 µL	500	0.1	60%	432	Normal
Positive	2 µL	500	0.05	60%	470	Normal
Positive	2 µL	1000	0.1	60%	342	Short
Positive	2 µL	1000	0.05	60%	367	Short
Positive	2 µL	50	0.1	70%	257	Very long
Positive	2 µL	50	0.05	70%	276	Very long
Positive	2 µL	250	0.1	70%	237	Long
Positive	2 µL	250	0.05	70%	258	Long
Positive	2 µL	500	0.1	70%	213	Normal
Positive	2 µL	500	0.05	70%	224	Normal
Positive	2 µL	1000	0.05	70%	176	Short
Positive	2 µL	1000	0.1	70%	173	Short
Positive	2 µL	50	0.1	80%	135	Very long
Positive	2 µL	50	0.05	80%	141	Very long
Positive	2 µL	250	0.1	80%	129	Long
Positive	2 µL	250	0.05	80%	134	Long
Positive	2 µL	500	0.1	80%	119	Normal
Positive	2 µL	500	0.05	80%	119	Normal
Positive	2 µL	1000	0.1	80%	100	Short
Positive	2 µL	1000	0.05	80%	100	Short

Table 4. Coverage, time for analysis, and confidence under different parameter settings in the negative ion mode

Polarity	Volume	Min Threshold	Mass slice widh	Identification cut off	Number of hits	Time for analysis
Negative	5 µL	50	0.05	60%	343	Very long
Negative	5 µL	50	0.1	60%	307	Very long
Negative	5 µL	250	0.05	60%	295	Long
Negative	5 µL	250	0.1	60%	273	Long
Negative	5 µL	500	0.05	60%	239	Normal
Negative	5 µL	500	0.1	60%	231	Normal
Negative	5 µL	1000	0.05	60%	177	Short
Negative	5 µL	1000	0.1	60%	172	Short
Negative	5 µL	50	0.05	70%	217	Very long
Negative	5 µL	50	0.1	70%	198	Very long
Negative	5 µL	250	0.05	70%	202	Long
Negative	5 µL	250	0.1	70%	189	Long
Negative	5 µL	500	0.05	70%	174	Normal
Negative	5 µL	500	0.1	70%	168	Normal
Negative	5 µL	1000	0.05	70%	134	Short
Negative	5 µL	1000	0.1	70%	132	Short
Negative	5 µL	50	0.05	80%	95	Very long
Negative	5 µL	50	0.1	80%	94	Very long
Negative	5 µL	250	0.05	80%	91	Long
Negative	5 µL	250	0.1	80%	91	Long
Negative	5 µL	500	0.05	80%	87	Normal
Negative	5 µL	500	0.1	80%	87	Normal
Negative	5 µL	1000	0.05	80%	76	Short
Negative	5 µL	1000	0.1	80%	75	Short

Each row in **Tables 3 and 4** is color-coded according to their identification confidence. Dark orange and light orange coded rows denote high confidence, while blue coded rows denote lower confidence and may contain more false positive results. Grey-coded rows represent little increase in the number of hits but indicate significantly increased analysis time. For context, the time for analysis of a single sample is presented in **Table 5**. Changes in the "Min Threshold" parameter can dramatically affect the analysis time but have less affect the number of hits. The identification cut-off percentage has a significant impact on the number of hits. And the "mass slice width" affects the number of hits only for data with low confidence.

Table 5. Typical times for analysis of 1 sample at different threshold values

Threshold	Analysis time (min)	Descriptor
50	12	Very long
250	1.7	Long
500	1.4	Normal
1000	1.2	Short

As shown in **Tables 3 and 4**, a balance must be struck between the analysis time, coverage, and confidence. In the positive ion mode, a minimum threshold of >50 and <250 appears to give the best coverage. Due to the high sensitivity and the high signal-to-noise ratio of the data acquired on the ZenoTOF 7600 system, a value of 250 cps was chosen to give the best coverage using an "identification cutoff" value of 70%. As seen in the tables, however, small changes can significantly affect the overall data, so it is recommended that

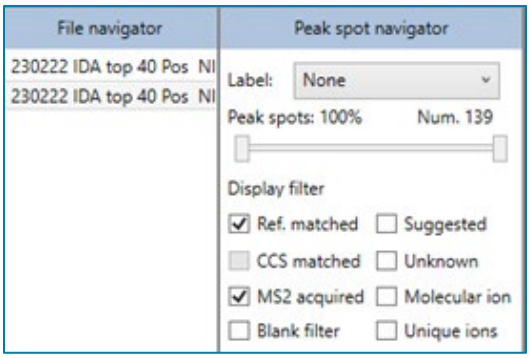

Figure 8. Peak spot navigator. MS2 acquired and Ref. matched are selected to filter the metabolites with the most confident annotation.

these software processing parameters be a starting point from which adjustments can be made to generate the best data for a given data set. For example, for data processed from a large cohort study. In that case, increasing the "minimum threshold" to 500-1000 cps may be reasonable to decrease the time needed for analysis. As expected, the data also indicate that a higher sample load (i.e., higher analyte concentration) generates higher-quality data.

The confidence level in the processed data can be adjusted in the "Peak spot navigator." Clicking on the "Ref. matched" display option will only show compounds with a matching MS/MS reference spectrum in the compound library (**Figure 8)**.

Although MS-DIAL's MS/MS fragment library algorithm matches acquired data with a high degree of confidence, it is recommended to manually validate the metabolite annotations from the right bottom panel (Identification) and change the annotation, if necessary, as shown in **Figure 9.**

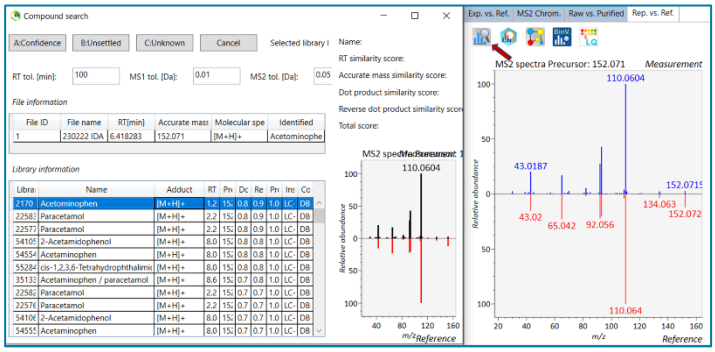

Figure 9. Identification panel. Annotated metabolites can be reviewed from this panel. If the annotation is not correct, it can be easily changed using the MS/MS look-up button (red arrow)

Conclusions

- The speed and sensitivity of the ZenoTOF 7600 system enable the generation of reproducible and high-quality data for untargeted metabolomic analysis
- In MS-DIAL 4.92 software, the “Minimum threshold” value of 250 cps and the “Identification cut off” value of 70% are good initial settings for untargeted metabolomics data acquired on the ZenoTOF 7600 system
- For extensive cohort studies, or in situations where it is necessary to minimize the overall analysis time, the “Minimum threshold” parameter setting value can be increased up to 1000 cps
- The number of analytes identified in a sample is proportional to the concentration of the sample; higher sample volume and concentration are recommended

References

1. Baker, PRS and Proos R. Untargeted data-dependent acquisition (DDA) metabolomics analysis using the ZenoTOF 7600 system. SCIEX technical note, RUO MKT-02-15367-A
2. Sayers R, Proos R, and PRS Baker. Analysis of untargeted metabolomics data from an untargeted Zeno data-dependent acquisition (DDA) workflow using SCIEX OS software. SCIEX technical note, RUO-MKT-02-15619-A
3. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M. Nat Methods. 2015; 12(6):523-6. Doi: 10.1038/nmeth.3393
4. Alseekh, S et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods. 2021; 18(7):747-756. Doi 10.1038/s41592-021-01197-1

Headquarters
500 Old Connecticut Path | Framingham, MA 01701
USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please all the division headquarters or refer to our website at
sciex.com/offices