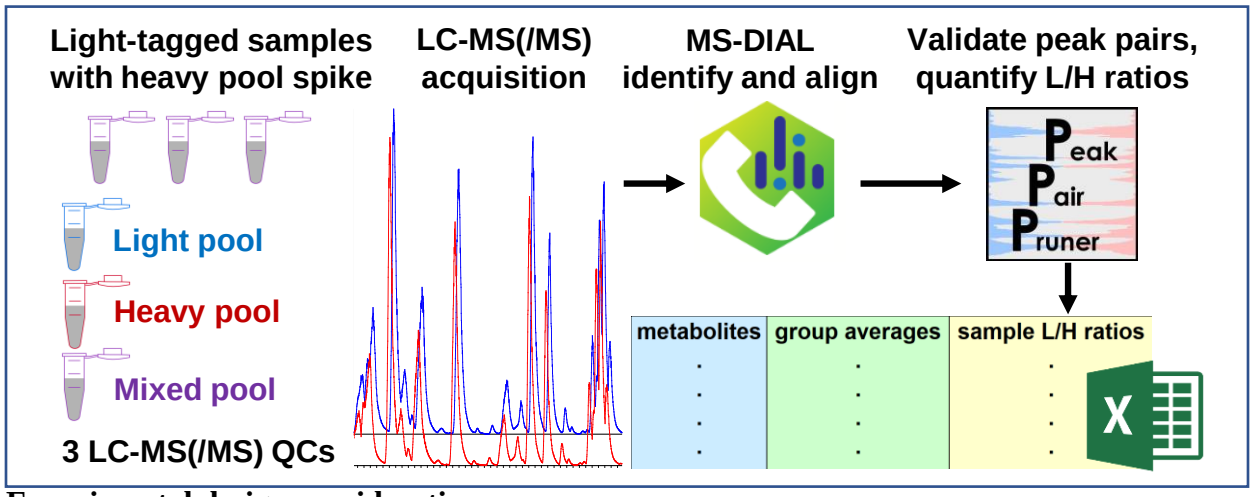
QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data



Experimental design considerations:

Past isotopic labeling data may or may not be suitable for this workflow based on the experimental design used. If evaluating data compatibility or planning a new isotopic labeling experiment with PPP in mind, the following samples with LC-MS(/MS) chromatograms are required:

- A blank. A reagent blank is best, such as a 1:1 mixture of a light reagent blank and a heavy reagent blank, but even a mobile phase blank will do. Triplicate or higher sampling is recommended. Blanks should run in series with your samples and QCs.
- 3 pool QCs: light-tagged, heavy-tagged, and mixed. You should pool equal aliquots of all samples prior to derivatization and from this common pool create a light-tagged pool and heavy-tagged pool. These two pools serve as essential PPP QCs. The third pool QC is a mixture of the first two pools at a known ratio. 1:1 mixing is recommended for most reliable quantification, but this can be increased to 2:1 or even 10:1 at high analyte concentrations if the scarcity of the heavy tag demands. Triplicate or higher sampling is recommended. It is also recommended that the heavy-tagged pool QC be analyzed in series with but before other QCs and samples, such as: Blanks, heavy pool, and then the rest.
- Light-tagged samples spiked with heavy pool. Any number of samples and replicates is permitted with triplicate or higher sampling recommended if instrument time allocation allows. First derivatize individual samples with your light tag and then spike in heavy-tagged pool at the same ratio as your mix pool QC.

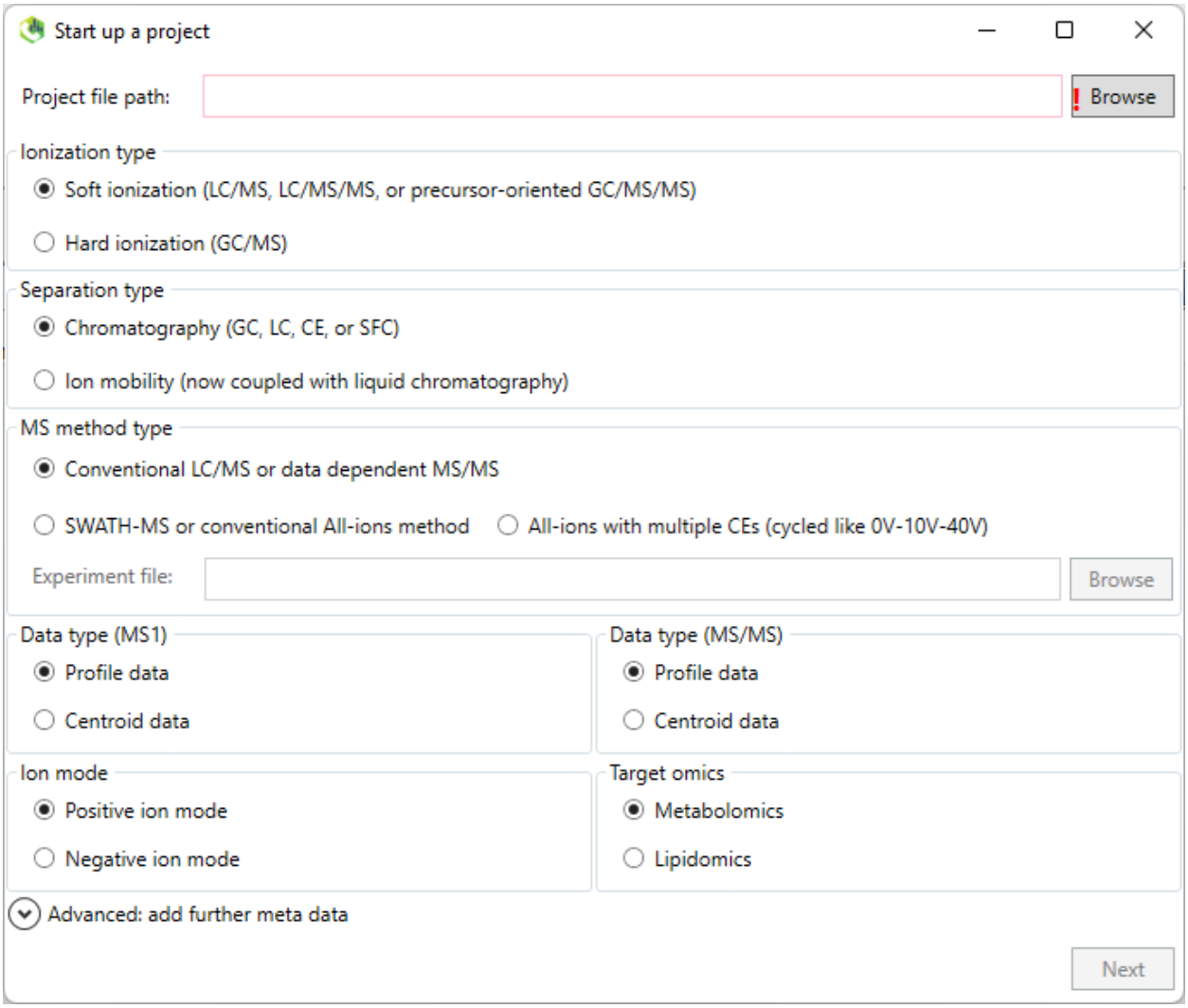
This document will discuss the PPP workflow with examples referencing the following dansylated[1] human plasma quantitation testing experiment with .ibf LC-FTMS datafiles on Github:

- Heavy reagent blank sampled in duplicate
- Light reagent blank sampled in duplicate
- Heavy, light, and 1:1 mix pool QCs sampled in triplicate
- Samples imitating a 5 individual cohort with 100-fold metabolite concentration spread:
    - 0010P: L/H 1:10 sampled in triplicate
    - 0050P: L/H 1:2 sampled in triplicate
    - 0100P: L/H 1:1 sampled in triplicate
    - 0150P: L/H 2:1 sampled in triplicate
    - 1000P: L/H 10:1 sampled in triplicate

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

MS-DIAL 4.9.2 Windows x64 peak alignment, identification, and isotopic pairing:

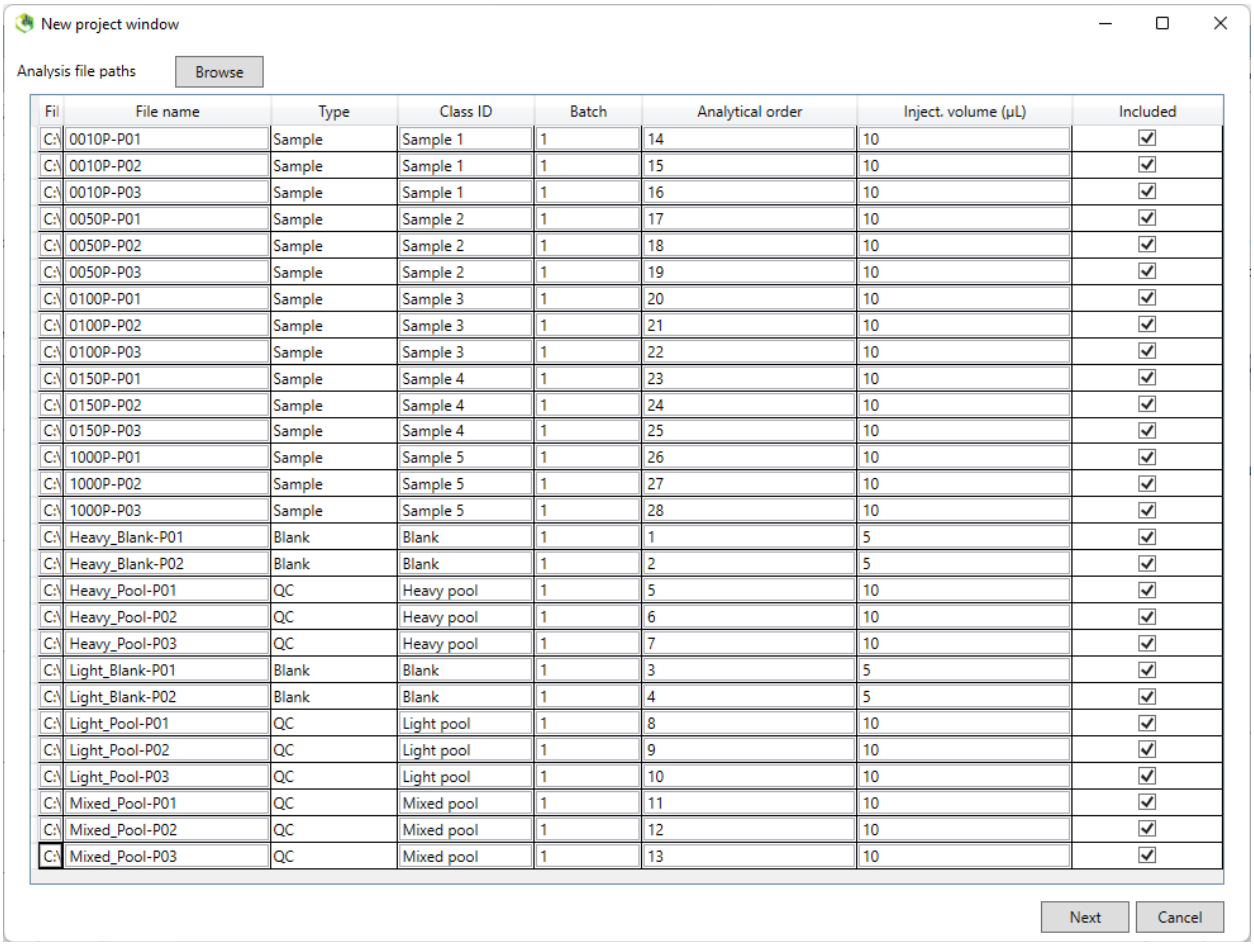
MS-DIAL[2] is a generalized metabolomics platform for LC-MS(/MS) and GC-MS analysis. A MS-DIAL project must be set up in a particular way if the goal is PPP peak validation and quantification. This document will walk through these particulars and briefly summarize the other parameters for the example human plasma dataset. To learn more about MS-DIAL's capabilities and the function of its parameters, refer to the detailed online tutorial. Our final data was reported utilizing MS-DIAL 4.9.2, but we have also tested it successfully using MS-DIAL 4.9.0 and MS-DIAL 5.1. For this dataset, the main difference between MS-DIAL 4.9.0 and MS-DIAL 4.9.2 or 5.1 is that conversion from .raw to .ibf is no longer required. In our experiments, we have so far found that continuing to use .ibf files is superior. It seems that there is some data loss by MS-DIAL when using .raw instead of first converting to .ibf. As such, you should convert your .raw files to .ibf files. MS-DIAL versions 4.9.0 and earlier come with a simple .ibf converter program that can be used to convert files for later versions. For the example data, .ibf files are provided.



Make a new folder for your MS-DIAL project to contain all of its related files and set this folder as the project path. When using an in-house identification library as in this example experiment, you must select Metabolomics beneath Target omics.

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

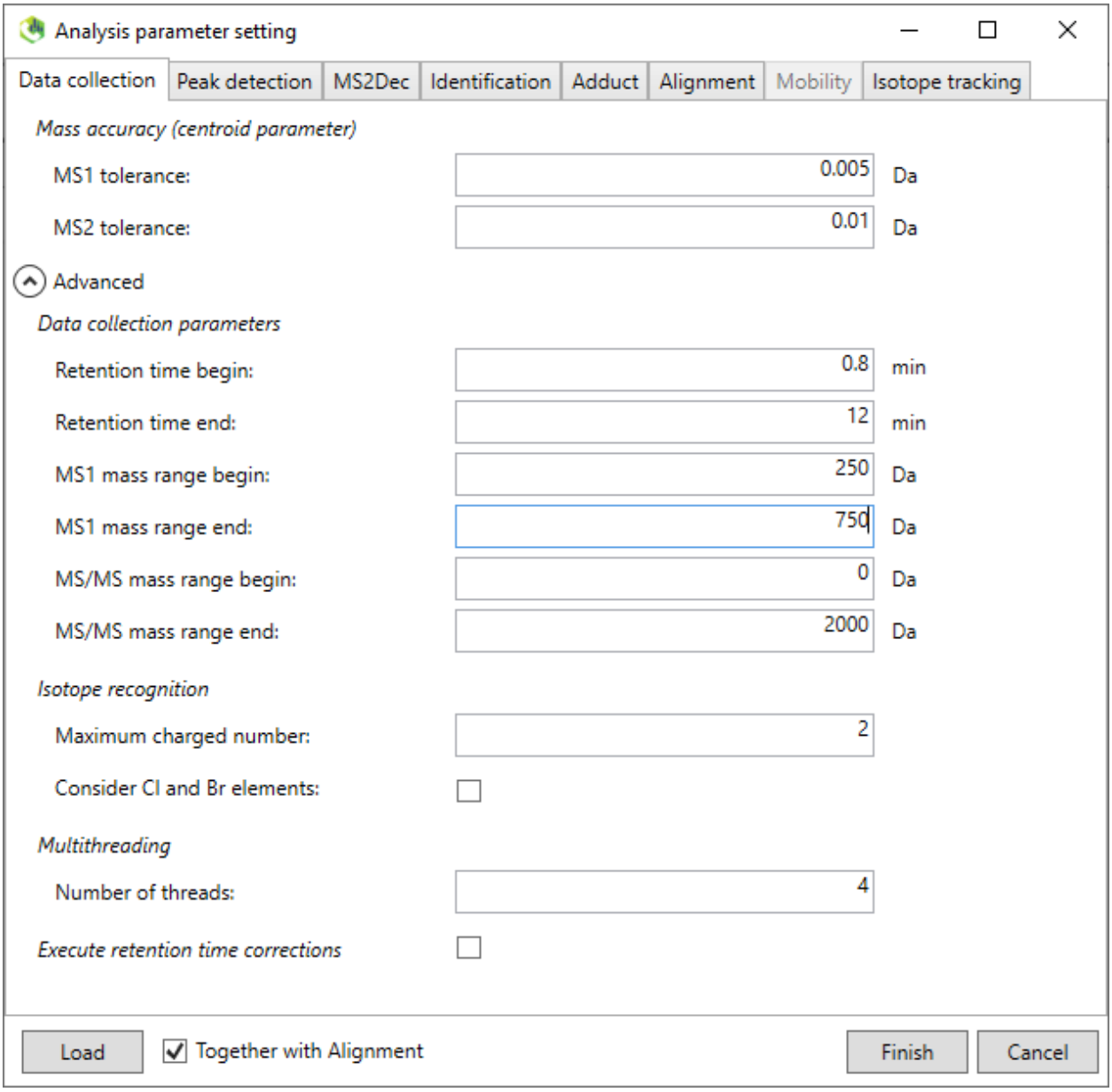
MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



Fil	File name	Type	Class ID	Batch	Analytical order	Inject. volume (µL)	Included
C:\	0010P-P01	Sample	Sample 1	1	14	10	☑
C:\	0010P-P02	Sample	Sample 1	1	15	10	☑
C:\	0010P-P03	Sample	Sample 1	1	16	10	☑
C:\	0050P-P01	Sample	Sample 2	1	17	10	☑
C:\	0050P-P02	Sample	Sample 2	1	18	10	☑
C:\	0050P-P03	Sample	Sample 2	1	19	10	☑
C:\	0100P-P01	Sample	Sample 3	1	20	10	☑
C:\	0100P-P02	Sample	Sample 3	1	21	10	☑
C:\	0100P-P03	Sample	Sample 3	1	22	10	☑
C:\	0150P-P01	Sample	Sample 4	1	23	10	☑
C:\	0150P-P02	Sample	Sample 4	1	24	10	☑
C:\	0150P-P03	Sample	Sample 4	1	25	10	☑
C:\	1000P-P01	Sample	Sample 5	1	26	10	☑
C:\	1000P-P02	Sample	Sample 5	1	27	10	☑
C:\	1000P-P03	Sample	Sample 5	1	28	10	☑
C:\	Heavy_Blank-P01	Blank	Blank	1	1	5	☑
C:\	Heavy_Blank-P02	Blank	Blank	1	2	5	☑
C:\	Heavy_Pool-P01	QC	Heavy pool	1	5	10	☑
C:\	Heavy_Pool-P02	QC	Heavy pool	1	6	10	☑
C:\	Heavy_Pool-P03	QC	Heavy pool	1	7	10	☑
C:\	Light_Blank-P01	Blank	Blank	1	3	5	☑
C:\	Light_Blank-P02	Blank	Blank	1	4	5	☑
C:\	Light_Pool-P01	QC	Light pool	1	8	10	☑
C:\	Light_Pool-P02	QC	Light pool	1	9	10	☑
C:\	Light_Pool-P03	QC	Light pool	1	10	10	☑
C:\	Mixed_Pool-P01	QC	Mixed pool	1	11	10	☑
C:\	Mixed_Pool-P02	QC	Mixed pool	1	12	10	☑
C:\	Mixed_Pool-P03	QC	Mixed pool	1	13	10	☑

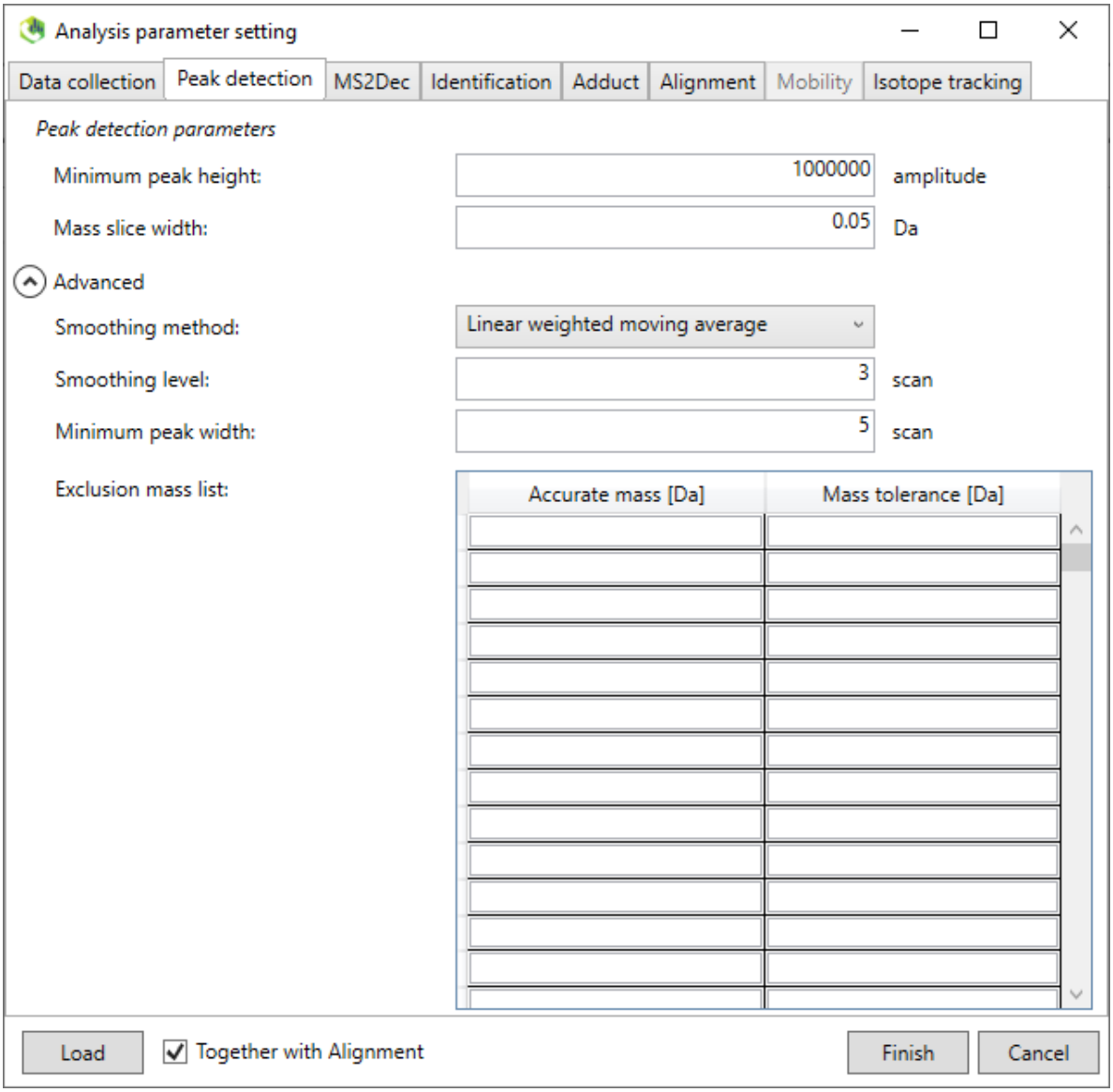
This is the most important window to double and triple check. A single character error on this page can cause great frustration later on. In the new project window, assign the appropriate Type for your samples, pooled QCs, and blanks. Providing the correct Class IDs is essential for downstream quantitation. Your light pool replicates must have the same Class ID (recommend copy-pasting one name), and this name must begin with either ‘Light’ or ‘light’. Your heavy pool replicates must also have the same class ID, and this name must begin with either ‘Heavy’ or ‘heavy’. Your mixed pool replicates must also have the same class ID, and this name must begin with either ‘Mix’ or ‘mix’. Regardless of blank strategy, your banks must have the same class ID, and this name must begin with either ‘Blank’ or ‘blank’. In this example dataset, two different blanks were used, a light reagent blank and a heavy reagent blank. Due to reagent scarcity, these two blanks were both diluted by a factor of 2 to reach sample volume. To account for this dilution, the injection volumes were entered such that the reagent blanks have half the injection volumes of the QCs and samples. The function of the Inject. volume column is to account for dilution variations across samples with the actual injection volume being potentially irrelevant. For your cohort samples, the Class ID field represents how you can group LC-MS(/MS) data. In this example, triplicate samplings of the 5 samples are grouped through the use of 5 sample Class IDs.

MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



Mass tolerances, retention time windows, and mass windows must be carefully considered based on your LC-MS(/MS) system and the samples analyzed. For this workflow, the maximum charge number should be the maximum charge number in your library. The number of threads refers to the number of processors your computer will dedicate to processing the data. More threads means faster processing but greater allocation of computer resources to MS-DIAL. Assigning too many threads can lead to freezing. The Load button can be used to load previously saved parameters. The parameter file for this example dataset is on Github. The parameter file will not automatically set your library file, alignment reference file, and isotope tracking parameters. To prevent lost time due to incorrect processing, all parameters should be double-checked even when loading a parameter file.

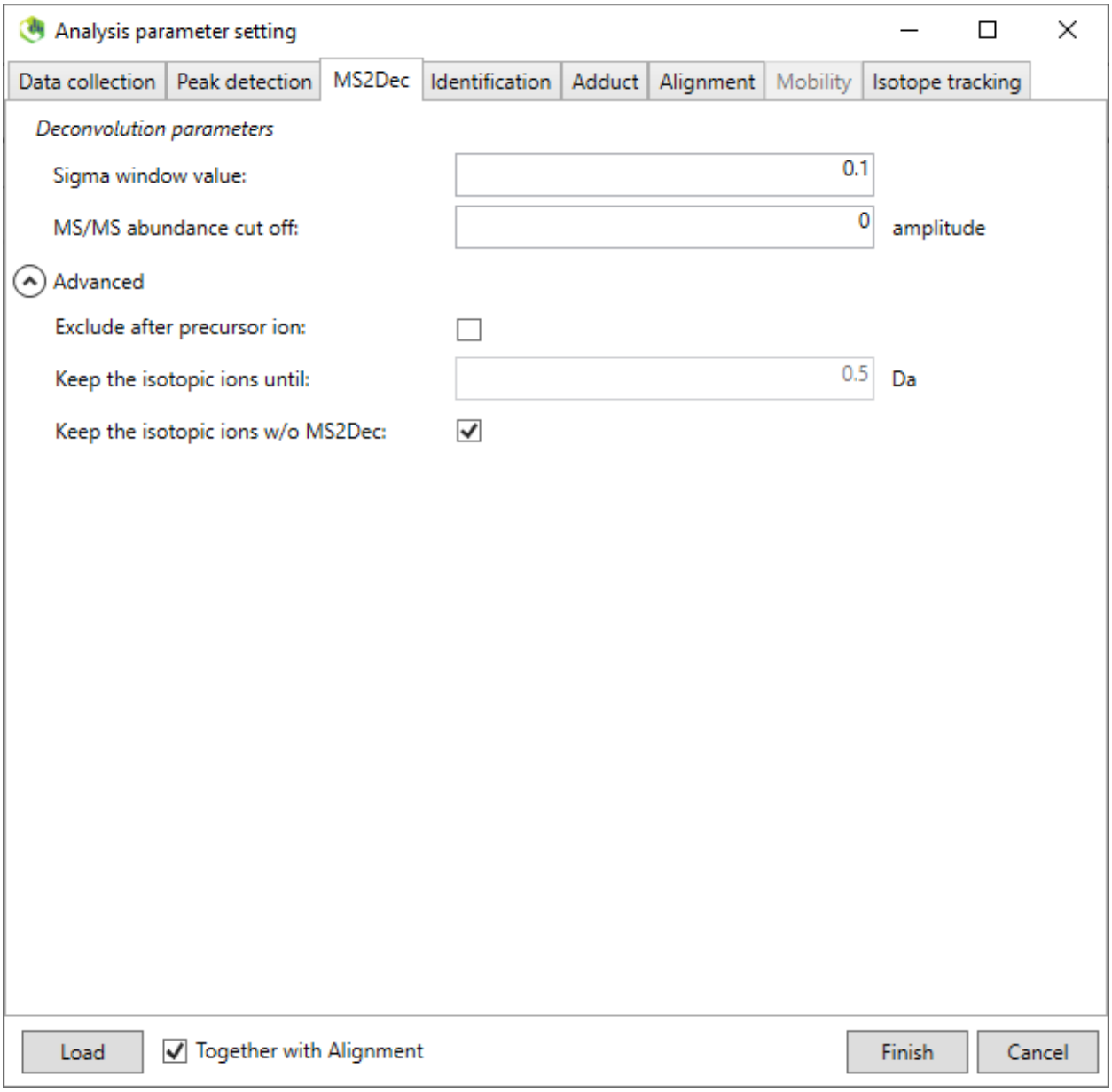
MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



The minimum peak height is an essential consideration in the PPP workflow. Peaks with intensities below this threshold will be eliminated from MS-DIAL's processing. Raising this value allows for faster data processing and reduces the probability that MS-DIAL will assign incorrect isotopic relationships between analytes and interfering peaks. Lowering this value increases the sensitivity of the workflow and widens the metabolome coverage. Regardless of tagging scheme, peak pairing will only be correct for a given analyte if its light-tagged M+1 peak is above the minimum peak intensity. In the example dataset, the highest analyte intensities are in the low-to-mid 10^9 range, and the background peaks are in the mid 10^5 range, making 10^6 a reasonable minimum. An exclusion mass list can be included to potentially improve isotopic peak pairing with a lower minimum intensity by excluding major background peaks. The other parameters refer to MS-DIAL's peak detection and smoothing algorithms. Reducing the smoothing level to 2 may yield better results if you are struggling with partially overlapping isomeric peaks.

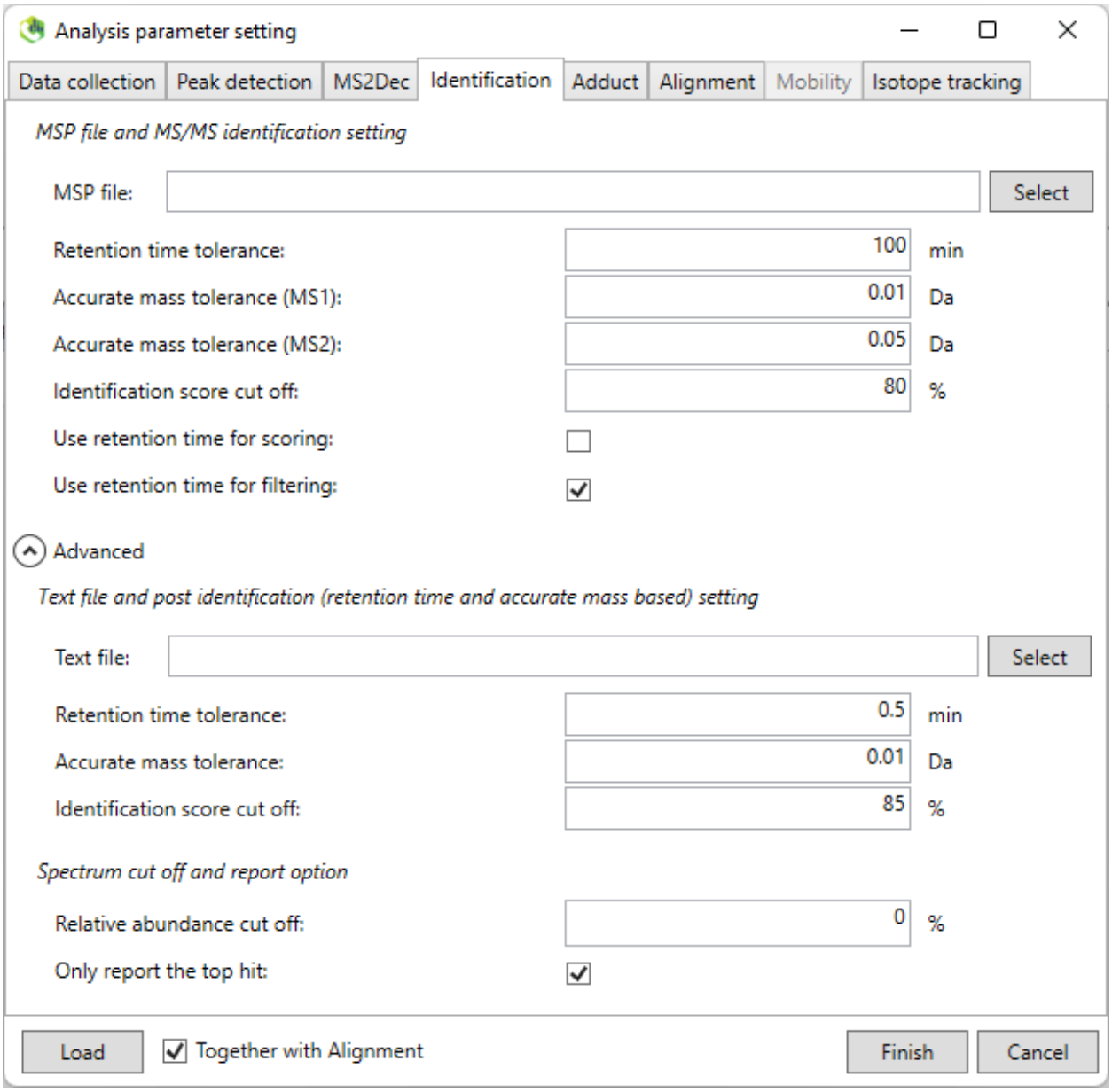
QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



Regardless of whether or not your LC-MS methods includes tandem mass spectrometry, the advanced MS2Deconvolution parameters must be assigned as above for downstream PPP analysis.

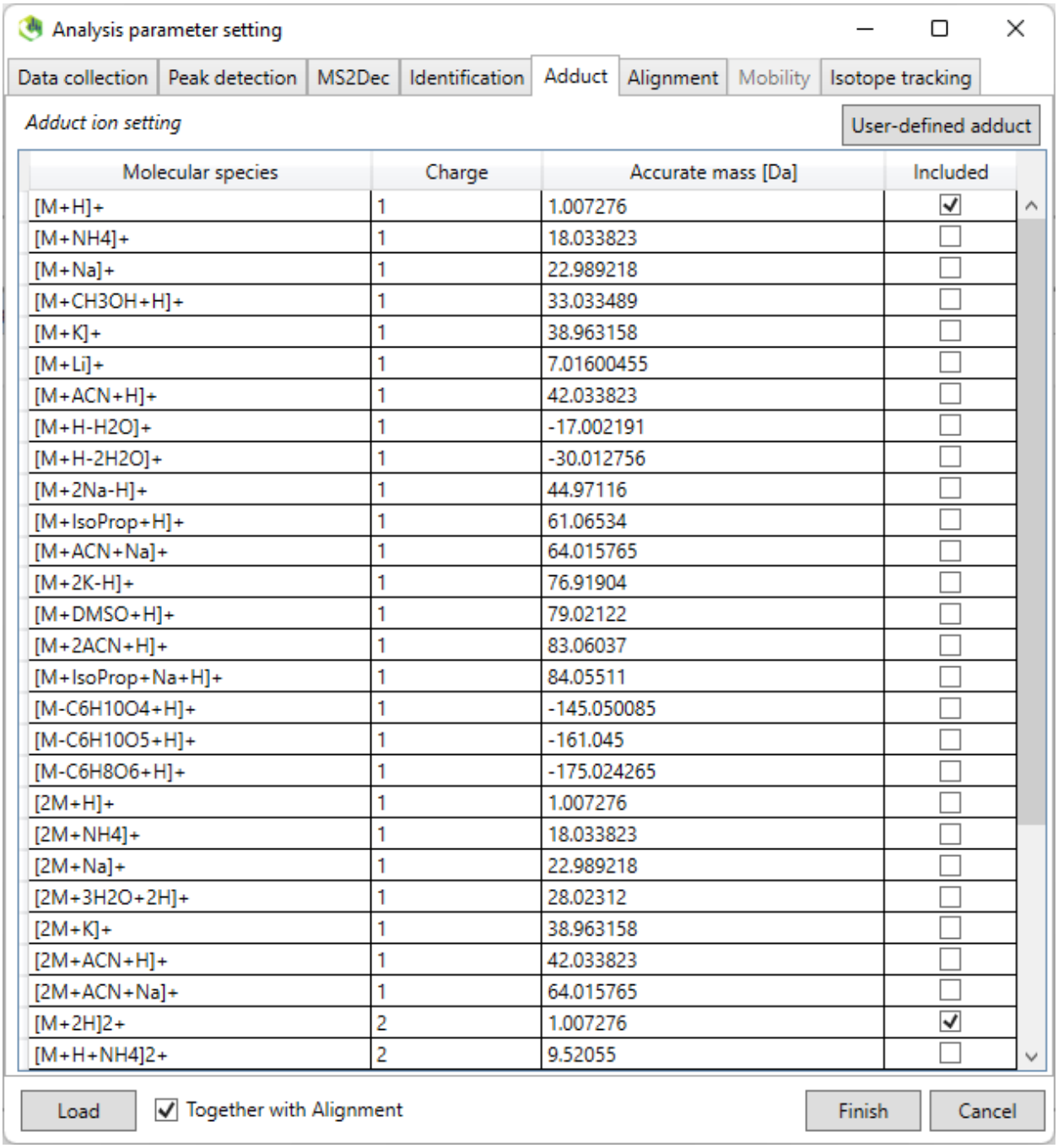
MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



In this example, the library used is a text file library with MS1, RT, and adduct type information. Select the provided DnsCl library on Github. MSP libraries can also be used containing fragmentation data for more confident identification.

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:

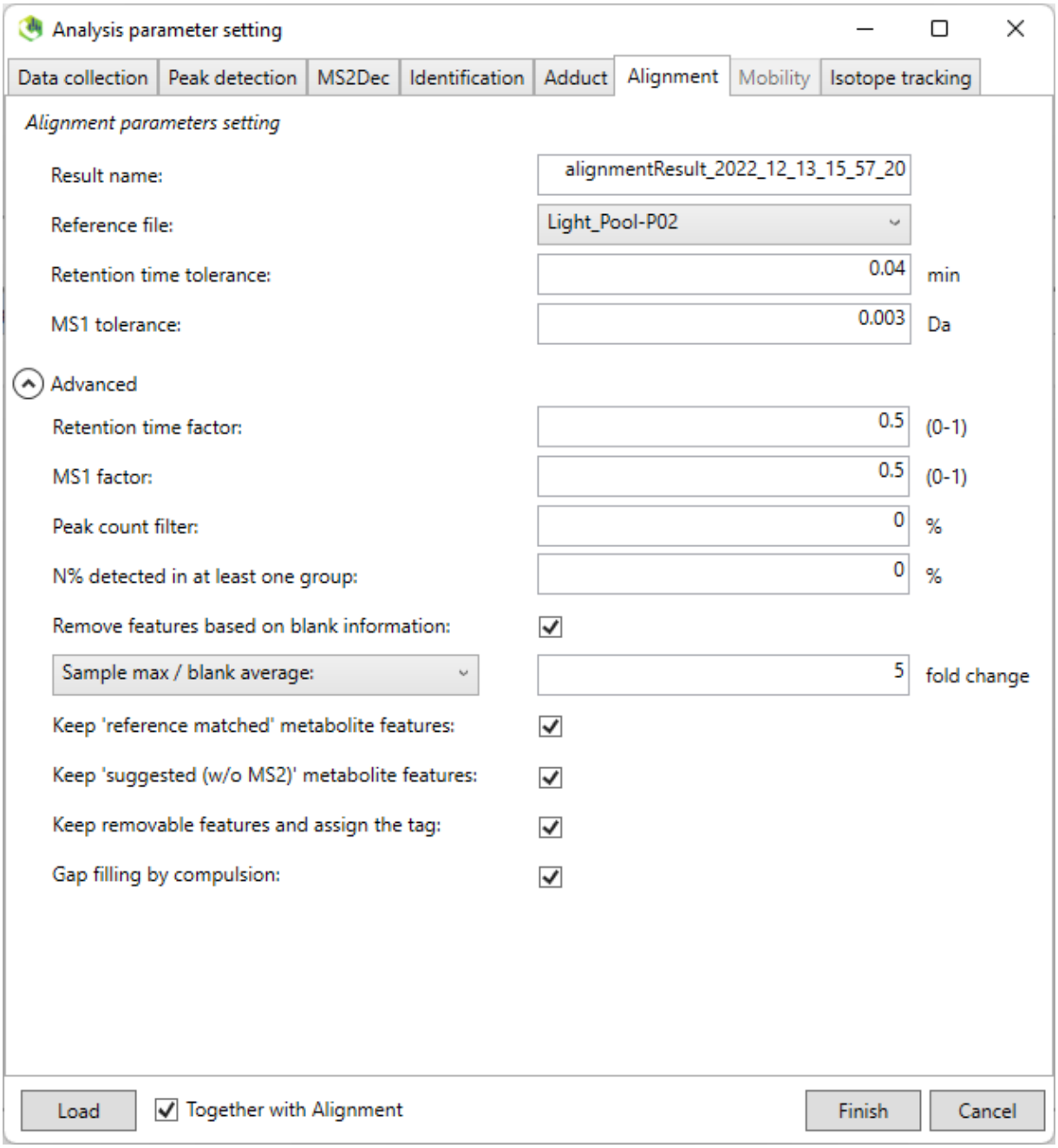


Molecular species	Charge	Accurate mass [Da]	Included
[M+H]+	1	1.007276	☑
[M+NH4]+	1	18.033823	☐
[M+Na]+	1	22.989218	☐
[M+CH3OH+H]+	1	33.033489	☐
[M+K]+	1	38.963158	☐
[M+Li]+	1	7.01600455	☐
[M+ACN+H]+	1	42.033823	☐
[M+H-H2O]+	1	-17.002191	☐
[M+H-2H2O]+	1	-30.012756	☐
[M+2Na-H]+	1	44.97116	☐
[M+IsoProp+H]+	1	61.06534	☐
[M+ACN+Na]+	1	64.015765	☐
[M+2K-H]+	1	76.91904	☐
[M+DMSO+H]+	1	79.02122	☐
[M+2ACN+H]+	1	83.06037	☐
[M+IsoProp+Na+H]+	1	84.05511	☐
[M-C6H10O4+H]+	1	-145.050085	☐
[M-C6H10O5+H]+	1	-161.045	☐
[M-C6H8O6+H]+	1	-175.024265	☐
[2M+H]+	1	1.007276	☐
[2M+NH4]+	1	18.033823	☐
[2M+Na]+	1	22.989218	☐
[2M+3H2O+2H]+	1	28.02312	☐
[2M+K]+	1	38.963158	☐
[2M+ACN+H]+	1	42.033823	☐
[2M+ACN+Na]+	1	64.015765	☐
[M+2H]2+	2	1.007276	☑
[M+H+NH4]2+	2	9.52055	☐

Select adducts relevant to your library file. In this example, the library contains [M+H]+ and [M+2H]2+ adduct modes.

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

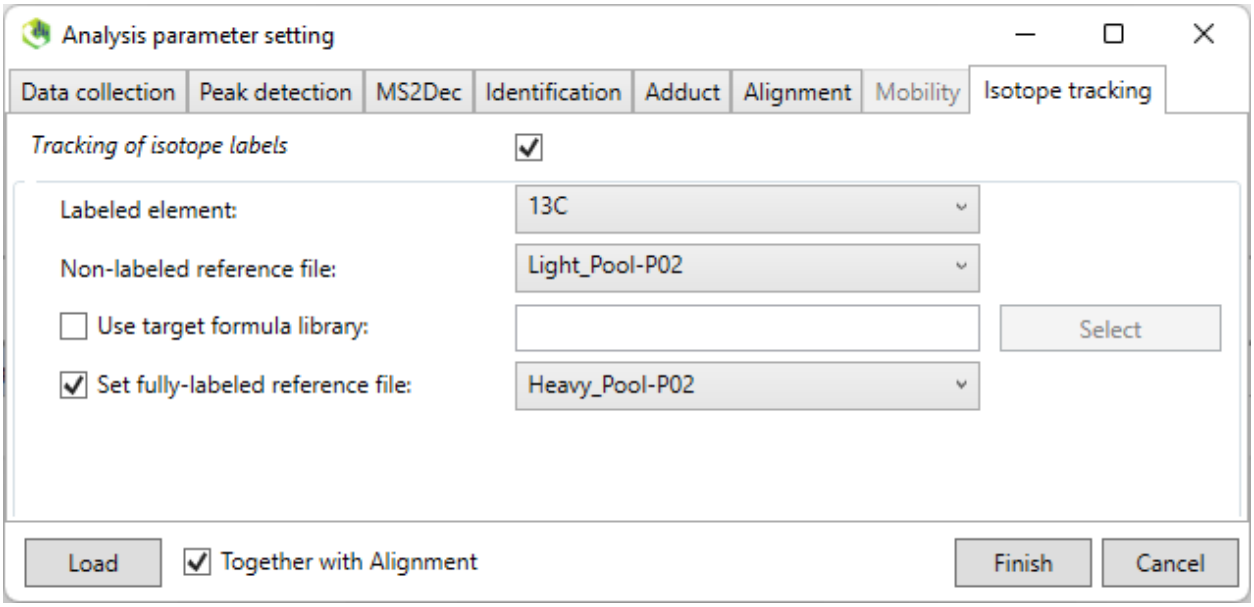
MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



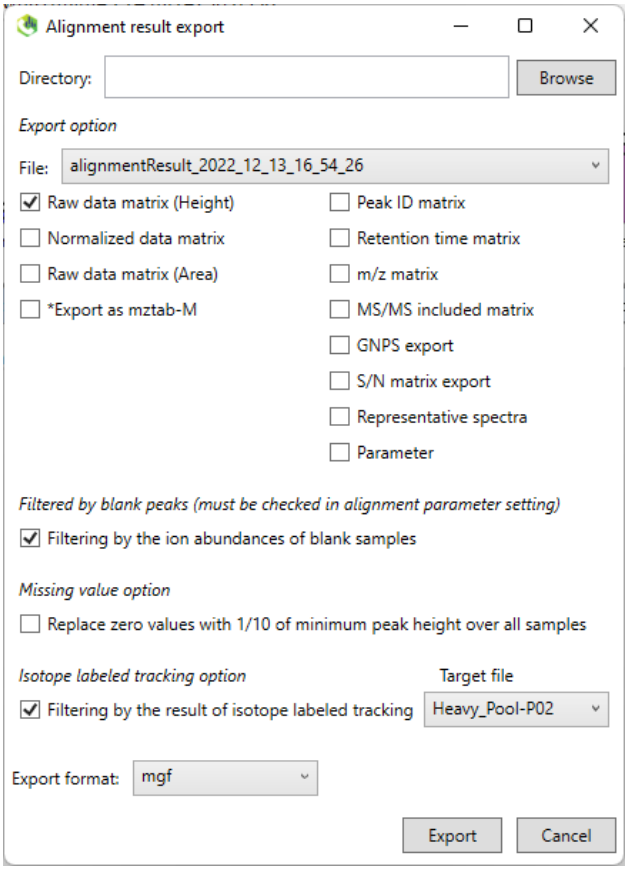
Several alignment parameters are essential for the PPP workflow. Your alignment reference file should be your light pool QC. The retention time tolerance and MS1 tolerance determine the windows within which peaks are aligned between your chromatograms. It is important to make sure your LC system is well equilibrated before and after each injection so that you can make the retention time tolerance small. High mass accuracy allows for a small MS1 tolerance. Reducing these tolerances, like raising the minimum ion intensity, decreases the probability that MS-DIAL will assign false isotopic relationships. Raising the minimum ion intensity and less complicated sample matrixes allows these two tolerances to be relaxed.

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

MS-DIAL 4.92 Windows x64 peak alignment, identification, and isotopic pairing cont.:



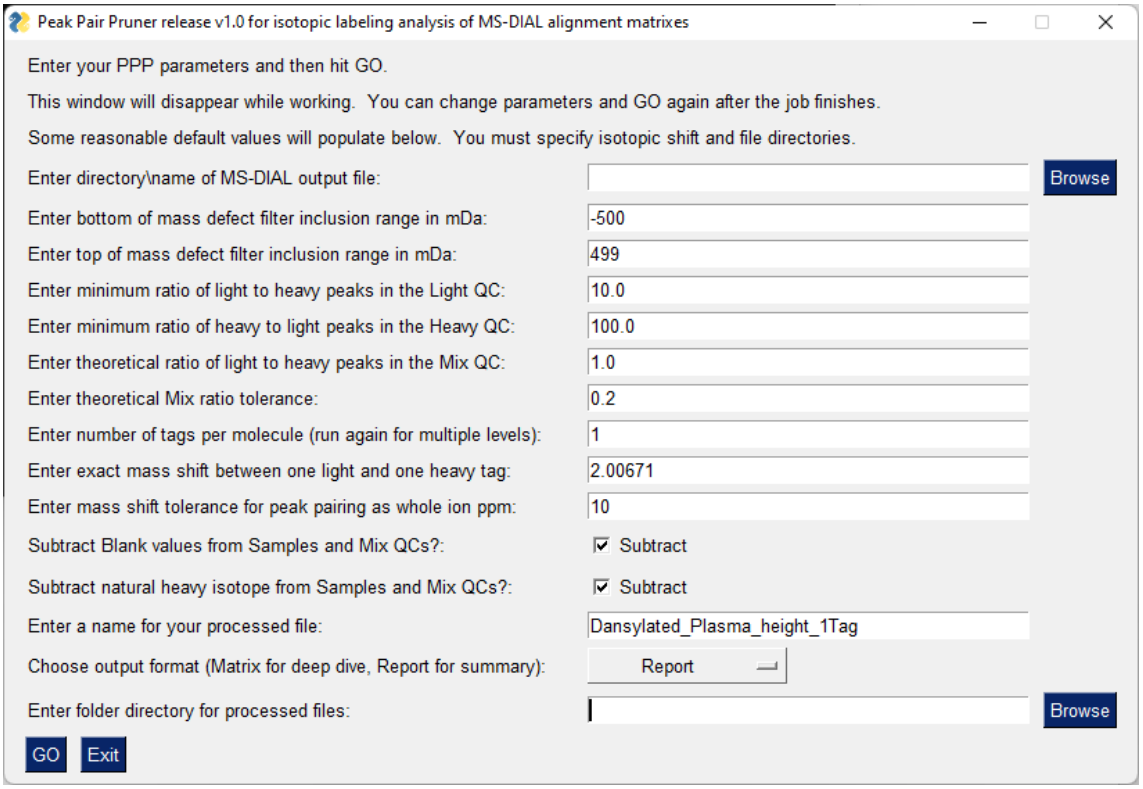
Enable tracking of isotope labels and set the labeled element for your tagging scheme. Dansylation relies on carbon-13.[1] Your non-labeled reference file should be the light pool QC. Enable the fully-labeled reference file option and set the file as your heavy pool QC. Double-check all of your parameters and click Finish to begin processing.



Export your alignment results as a raw data matrix. You can export an average peak height raw data matrix or a peak area raw data matrix, but do not select both options for a single export file. Enable filtering by the ion abundances of blank samples if you want MS-DIAL to eliminate background peaks. Filter by the result of isotope labeled tracking in your heavy pool QC. In our experiments with FTMS data acquisition and virtually no RT shift between light and heavy peaks, we have had better results with average height than with peak area. Simultaneous acquisition of both light and heavy peaks by the orbitrap allows for peak height to supersede peak area for most accurate quantification.

QZhangLab: PPP isotopic labeling metabolomics workflow instructions with example data

PPP validation and correction of peak pairs:



Run the PPP .exe and two windows will appear. Keep both open. The parameter window is shown above with the example data values filled in. Select your MS-DIAL alignment matrix and select an output directory. Optional mass defect filtering can be applied based on your experiment. The current mass defect window of [-500 mDa, 499 mDa] is all inclusive and so effectively disables this filter. If you are analyzing a particular class of analyte, the mass defect window may be useful for focusing your results. Set a minimum acceptable L/H ratio in your light pool QC. This value should be large in your light pool QC as there are no heavy-tagged molecules. However, your light pool analytes will still have natural heavy isotope abundance which may overlap with your heavy tag's m/z values. Enter a minimum acceptable H/L ratio in your heavy pool QC. Likewise, this value should be large but does not suffer from overlap like the light pool QC. Enter your theoretical mix ratio from your experimental design and an acceptable tolerance for peak pairs. 1.0 and 0.2 for these values indicates that a peak pair with L/H in the range [0.8, 1.2] will pass validation by the mix pool QC. Enter the number of tags per analyte molecule. If you have multiple tagging levels across different analytes, simply run more than once with different tag number values. After running once, the window will return with your values saved, making this easy to do. For this example data, you should run with 1 tag and again with 2 tags. Enter the exact mass shift between your light and heavy tags, 2.00671 Da for 2 carbon-13. Enter the mass shift tolerance for checking peak pair mass shifts. Optionally subtract background peak pair values based on the blank. Optionally subtract isotopic overlap (natural heavy abundance) based on the light pool QC. Enter a name for your file. Most applications will benefit from the Report export format, but there is a modified matrix format as well for trouble-shooting.

PPP validation and correction of peak pairs (cont.):

```
Working...


Initial isotopic matches from MS-DIAL:
Identified sets: 98
Unknown sets: 701

Mass checks performed with provided tolerances:
Mass defect filter inclusion range bottom: -500 mDa
Mass defect filter inclusion range top: 499 mDa
Number of tags per molecule: 1
Exact mass shift between one light and one heavy tag: 2.00671 Da
Mass shift tolerance as whole ion ppm: 10.0 ppm
Identified sets: 96
Unknown sets: 657

Isotopic ratio checks performed with provided ratios parameters:
Minimum Light QC light:heavy ratio: 10.0
Minimum Heavy QC heavy:light ratio: 100.0
Theoretical Mix QC light:heavy ratio: 1.0
Theoretical Mix ratio tolerance: 0.2
Identified sets: 84
Unknown sets: 367


Finished.
You may change parameters and GO again.
```

The second PPP window will show you the progress of data processing (usually very fast) and gives a breakdown of where peak pairs are being eliminated if they fail validations. In this example, 98 identified peak pairs from MS-DIAL had isotopic relationships conforming to 1 dansyl chloride tag (M+0 and M+2). Two were eliminated by the ppm mass shift check. Twelve were eliminated in the QC checks. These twelve turn out to be doubly-tagged molecules (M+0 and M+4), which pass the validation checks when rerunning with 2 as the number of tags per molecule.

Troubleshooting:

If you have dramatically fewer validated peak pairs after PPP processing than you expected or than you had identified in MS-DIAL, you should first check the PPP companion window (above). This will tell you where the peaks were eliminated. If the initial isotopic matches from MS-DIAL is too low, then there is a pairing issue in MS-DIAL, which could be related to your alignment tolerances being too strict or your minimum peak height being too high. It could also indicate a chromatographic problem. If there are numerous losses in the mass check steps, make sure you have correctly calculated the mass shift between your light and heavy tag with as many significant figures as are available. Check that the mass defect filtering employed is not too restrictive for your target analytes. If mass accuracy is low, you may need to relax the ppm tolerance. If losses are great in the QC ratio checks, there may be an experimental issue such as differing reactivities and yields between the light and heavy tags or evaporation leading to biased mixing/spiking. Relaxing all tolerances to extremes and looking at the Matrix output format can help diagnose the exact problem. Losses can also be caused by incorrect peak pairing in MS-DIAL, which is not easily diagnosable in PPP. Check the problematic peak pairs in MS-DIAL and assess whether or not MS-DIAL is pairing them correctly or potentially not at all.

References

(1) Guo, K.; Li, L. Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal Chem* **2009**, *81* (10), 3919-3932.
(2) 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* (6), 523-526.