30 November 2020
WHONDRS Data Package
https://whondrs.pnnl.gov
whondrs@pnnl.gov

Instructions for Analyzing FT-ICR-MS Data

Required Programs and Files

Formularity (https://omics.pnl.gov/software/formularity)
- Database file (.bin)
- Requires Windows (not compatible with MacOS/Linux)

R (https://www.r-project.org/)
Provided calibration file (.ref)
Provided R scripts

Procedure Overview

This protocol will outline how to generate an FT-ICR-MS data report. In brief, the provided *.xml* files contain peak lists which were generated using Bruker Daltonik's Data Analysis v4.2 software (Bruker Corp, Billerica, MA). These peak lists will be loaded into Formularity, a program capable of calibrating peaks, combining samples, and assigning chemical formulas, to create an unprocessed ICR report and an accompanying *.log* file. In order to detect samples that may have calibrated poorly, the provided "Processing_Formularity_Log.R" script can be used to generate a list of suspicious samples.

Detailed Procedure

1. Download Formularity from https://omics.pnl.gov/software/formularity. While on the website, be sure to also download the "CIA_DB_2016_11_21.bin" database file.
    a. Downloading Formularity will also provide documentation and some example data.
    b. For further details regarding this program, please refer to *Tolić et al, 2017 - Anal. Chem.*
2. After downloading Formularity, ensure that the settings mirror those in Figure 1 (also listed below)
    a. Under the Spectra files section:
        i. Charge: 1
        ii. Ionization: proton_detachment
    b. Under the Calibration section, drag and drop the .ref file into the box specified "Drop calibration file" and keep the default settings (listed below if changed).
        i. Select "linear" from the drop-down menu
        ii. Start tolerance, ppm: 5
            1. If this start tolerance yields many poorly calibrated samples or gives many errors, this can be adjusted to 8 ppm. This should not go higher.
        iii. End tolerance, ppm: 0.5
        iv. Rel. factor: 1000000

c. Under the peak filters section, ensure:
    i. Min S/N: 7
    ii. Min rel. abund.: 0
    iii. Max rel. abund.: 1
d. Under the "CIA formula finder" tab:
    i. Drag and drop the .bin file into the box specified "Drop DB files". Settings should be default
    ii. "Alignment" needs to be checked
    iii. Alignment tolerance, ppm: 0.5
    iv. Formula tolerance, ppm: 0.5
    v. DB mass limit: 500
    vi. Formula score: min(N+S+P) & The lowest error
    vii. Max relationship gaps: 2
    viii. Error: AMU – 0.00002
    ix. Within the "Formula building blocks" box, ensure that CH2, H2, and O are checked.
    x. User-defined filter: O>0 AND (N+S+P)<6 AND S<3 AND P<2
    xi. Ensure that "Use relationship", "Use formula filters", and "CIA" (top right in the Spectra Files section) are all checked.

3. With the correct settings, you can now drag and drop the provided .xml files into the box labeled "Drop Spectra Files" in the upper right of Formularity.
    a. The box will turn red and Formularity will begin generating the report, which can take up to 30-45 minutes depending on sample number, and the program might seem unresponsive.
4. Once the program finishes running, two files will be generated (Report.csv and a log file) in the folder containing the .xml files.
    a. We recommend renaming the report to match your dataset name; this will prevent it from being overwritten if you run Formularity again.
5. Next, edit the "Processing_Formularity_Log.R" script to include your dataset name and directory containing the log file and run it through either R, RStudio, or command line.
    a. This will generate two files: one with all calibration results and another specifying poorly calibrated samples. **We recommend removing these samples.**
    b. Note: If you are analyzing WHONDRS 48 Hour Diel Cycling Study at HJ Andrews Experimental Forest Watershed 1 (WS1) data (doi:10.15485/1509695), this calibration check will fail.
6. You now have a finalized FT-ICR-MS Report! We recommend using the R package "ftmsRanalysis" (https://github.com/EMSL-Computing/ftmsRanalysis) to further process the data (i.e., filter peaks by mass, calculate molecular properties, classify compounds, etc.).

Note about sample names

The sample names for the FT-ICR data might look a little different than those found in the geochemistry. Specifically, files will be appended with "p" and then some variable number (i.e., p08, p1, etc.). These values correspond to the "ion accumulation time" (IAT) at which a sample was collected. The IAT is a key parameter in collecting FT-ICR data and can have significant effects (see *Cao et al, 2016 – Anal. Chem.*). These can be analyzed separately or together, and similar samples can even be merged after alignment, but these could alter data interpretations.

In limited cases, there might be replicates for some samples. These will be denoted by either "rep1" or "rep2". Unlike the IAT differences, these are functionally identical meaning they can be combined without any additional considerations (or analyzed separately based upon use case).

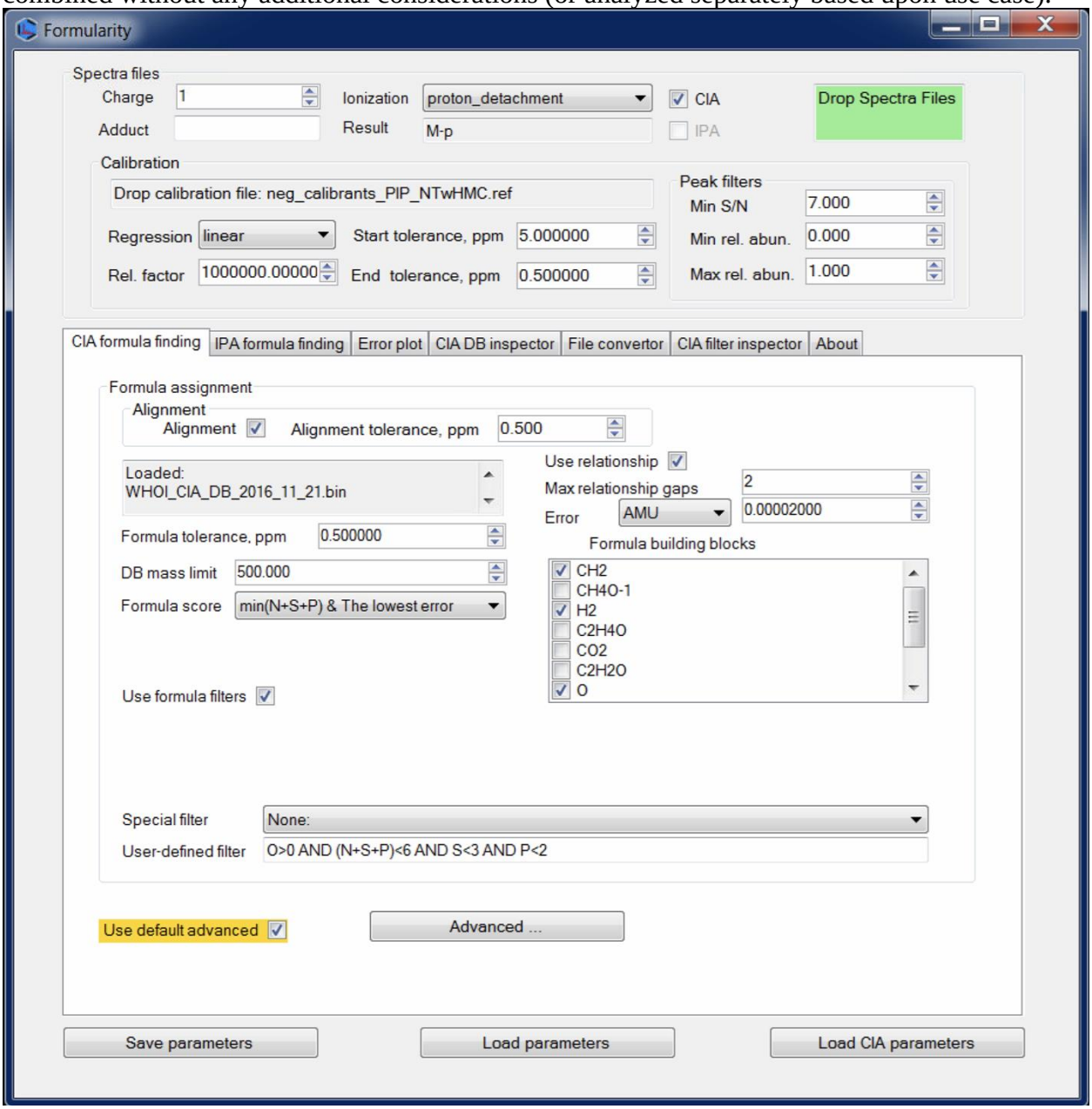

Figure 1: Default Formularity settings.