A FTICR Data Analysis Tutorial

Before we can analyze some of these data, we need a couple programs and files:

* All of the files (including this walkthrough) can be found on GitHub.
  + <https://github.com/danczakre/ICRTutorial>
* If you don’t have it installed already, please install RStudio.
  + <https://www.rstudio.com/products/rstudio/download/>
* Once RStudio is installed, please run the following commands:
  + *Please follow the commands found in the top directory on the associated GitHub page*
  + **Note:** Copying and pasting these commands may not work – they might need to be directly typed into RStudio. If you wish to copy and paste these commands rather than re-type them, there is a file named “install\_packages\_commands.txt” on the repository from which the commands can be copied.
* An extra program is necessary to learn alignment/calibration
  + Formularity: <https://omics.pnl.gov/software/formularity>
  + Be sure to download the **program** and the **database**

Once you have these files installed, you can follow along with the tutorial.

The primary goal of this tutorial is to demonstrate one potential avenue through which FTICR data you receive from EMSL can be analyzed. While many of these scripts were written to be broadly applicable, I can’t guarantee that they will work with all data types – only those FTICR reports generated by Formularity are currently acceptable. Furthermore, some of the figure generation steps are hard coded for the tutorial dataset. However, as EMSL transitions to other report formats, I will be creating another GitHub repository which will contain more broadly compatible analysis scripts/functions.

**Protocol:**

1. Open the folder containing Formularity and double-click the Formularity.exe icon
   1. Drag and drop the .ref file from GitHub into the calibration file box
   2. Drag and drop the database file from omics.pnnl.gov into the database file location
   3. Change the following settings
      1. CIA: Checked
      2. Ionization: proton\_detachment
      3. Adduct: ensure this is blank
      4. Under Calibration:
         1. Start tolerance, ppm: 5
         2. Regression: linear
      5. Under Formula Assignment:
         1. Alignment: Checked
         2. Alignment tolerance, ppm: 0.5
         3. Max relationship gaps: 2
      6. User-defined filter: O>0 AND (N+S+P)<6 AND S<3 AND P<2
   4. Drag and drop all .xml files into the green box labeled “Drop Spectra Files”
2. Run the **FTMS\_Analysis.R** script by double-clicking it, and then either clicking the “Source” button in the upper left-hand corner of the script editing window or pressing *Option+Command+R* on Macs or *Ctrl+Shift+Enter* on Windows.
   1. This script uses the R package upon which FREDA is based to take a Formularity report, filter peaks based upon mass range and isotopic signature, calculate stoichiometric measurements, and output a commonly formatted dataset.
   2. You should have two files named “Processed\_[Sample Name]\_Data.csv,” which contains the compound intensity by sample data, and “Processed\_[Sample Name]\_Mol.csv”, which contains the molecular characteristics.
3. Using these two files, you can run the **Generate\_MolProp\_Plot.R** script in the same way detailed in step 1.
   1. This script will generate a plot of some molecular formula property based upon your selection.
   2. This will output two separate plots: A boxplot that details the distributions of the provided property on a per-sample basis, and the average of many different properties across samples
4. You can also begin analyzing variations in the carbon character by running the **Generate\_VK\_Plot.R** script as detailed above.
   1. This will help you analyze the ‘molecular landscape’ of your samples, enabling you to comment upon the different types of carbon which are present.
   2. This will generate two different Van Krevelen plots: The first will plot a VK diagram for all of the molecular formula found across the dataset; the second will generate a plot for an individual sample that you specify.