Processing of Direct Infusion Mass Spectrometry (DIMS) data sets

Summary:

The purpose of this script is to take DIMS data files and arrange them in a data matrix (positive and negative mode separately) with columns as samples and rows as bins with summing up the intensities in each bin.

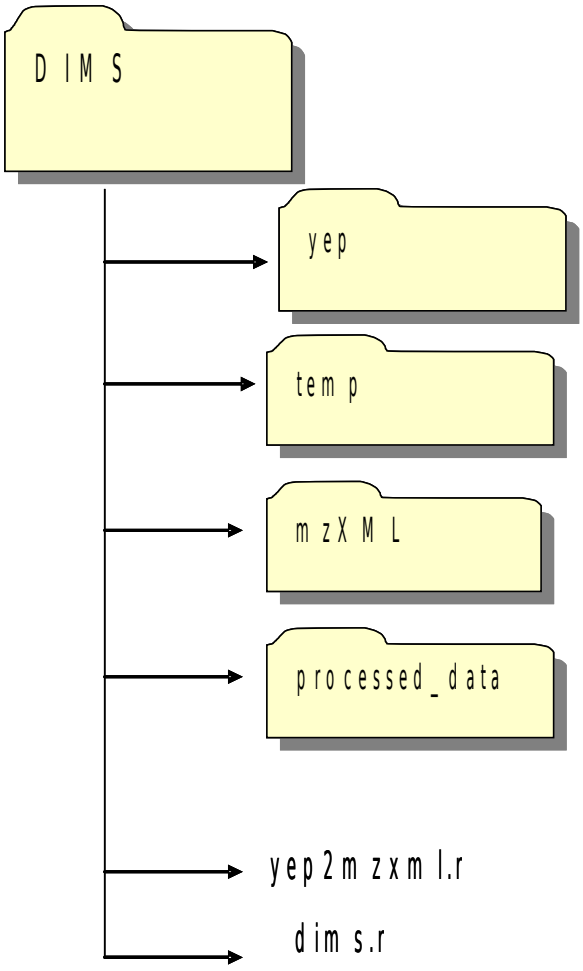
Step 1: Download the software CompassXport

Data from MS is in .yep files which are needed to be converted into mzxml file extensions to read it easily. An R script changes all the .yep files into .mzxml files using the software CompassXport that can be downloaded free from
http://www.ionsource.com/functional_reviews/CompassXport/CompassXport.htm

Please note that CompassXport only runs on the Microsoft Windows operating system.

Step 2: Directory Structure:

The directory structure for processing the DIMS data sets is given in Figure 1. The directory *yep* contains all the ".yep" files from the instrument. The directory *temp* is needed as a transit place to convert the files from extension ".yep" to ".mzxml" and put them in *mzXML* directory using the script "yep2mzxml.r". Data after processing using the script "dims.r"is placed in the *processed_data* directory.



Step 3: Running the script “yep2mzxml.r”

This script takes the “.yep” files from *yep* directory and convert them to “.mzxml” and place them in *mzXML* directory.

To run the script type:

Source(“yep2mzxml.r”)

yep2mzxml(inDir="yep",tempDir="temp",outDir="mzXML")

Step 4: Running the script “dims.r”:

The R function “dims.r” is created with the following input:

inDir: The directory containing the .mzxml files to be processed i.e. *mzXML*

outDir: The directory where output files will be saved i.e. *processes_data*

lowestMZ: lowest m/z ratio, default is 100.

higestMZ: highest m/z ratio, default is 2200.

deltaMZ: bin width, default is 1.

thresh: thresh hold for bin intensity, default is 0.

Setting the above input the script use R package XCMS to read the .mzxml files, it then bin the sum of intensities according to the bin width and create two out files pos.csv and neg.csv. Where neg.csv is the data collected in negative scan mode and pos.csv is the data collected in positive mode.

The first column consists of masses, first row as samples names and remaining entries are sum of intensities corresponding to each bin.

To run the script type:

source("dims.r")

dims(inDir="mzXML",outDir="processed_data", lowestMZ=?, higestMZ=?, deltaMZ=? , thresh=?)

Note that “?” need to be filled manually.