1 Introduction

Mass spectra, as well as their metadata, are considered as the raw measurement data and usually recorded in a vendor-specific binary format. During a measurement, the mass spectrometer captures internal heuristics which enable the instrument to adapt to sample properties like sample complexity or amount in near real time. Still, method parameters controlling these heuristics need to be set before the measurement. An optimal measurement result requires a carefully balanced set of parameters, but their complex interactions with each other make LC-MS method optimization a challenging task.

Here we present rawDiag, a platform-independent software tool implemented in R that supports LC-MS operators during the process of empirical method optimization. Our work builds on the ideas of the discontinued software rawMud [fastSci]. Our application is currently tailored towards spectral data acquired on Thermo Fisher Scientific instruments (raw format), with a particular focus on Orbitrap mass analyzers (ExactMax or Fusion instruments) and the PSI open proteomics file standard [1]. rawDiag is meant to run after mass spectrometry acquisition, optimally as an interactive R shiny application and produces a series of diagnostic plots.

2 Implementation

The entire software is implemented as an R [2] package providing full documentation and includes example data. A rough sketch of the package can be seen in Figure 1. All diagnostic plots are generated by R functions using the ggplot2 [3] graphical system, based on "The Grammar of Graphics" [4]. The package ships with an adapter function read.raw which returns an R data.frame object from the raw data input file. In its current implementation, the adapter functions default input method is set for reading Thermo Fisher Scientific raw files, using a C# programmed executable, based on the platform-independent New RawDiag::reader.Net assembly [2]. The package also ships with an S3 utility function as.rawDiag to coerce data from the PSI open proteomics file standard, e.g., by using the code snippet as.rawDiag(matrix(data.frame(...), nrow=5)) to fill in the by rawDiag used tidy data frame [8]. Figure 2 shows a comparison of both methods.

3 Application Example

The R code snippets below show the usage of the package optimization of an LC-MS/MS method.

```r
library(rawDiag)
data("W163763")
df <- W163763
```

Sample Data

This data was recorded to investigate the optimal number of MS2 scans between two consecutive MS1 scans on a Q-Exactive HF-X mass spectrometer injecting a commercial HeLa digest. The instrument data is available through http://www.bfabric.org [5] (workset W163763 sample S174020 or MassIVE M5V00882389). For the following demonstration we select three out of nine mass spectrometry runs.

```r
df <- df[grepl("[0-9]{6}\_S174020\", + dfEndline)]
```

Total Ion Chromatogram

First, we want to inspect the TIC or base peak chromatogram. With this plot we can see if the data was recorded properly and if the signal response of the sample is the same over the three injections.

```r
plot TicBasePeak(df, + method = "overlay")
```

Scan Frequency

As expected all three mass spectrometry methods result in different scan speed.

```r
plotScanFrequency(df, + method = "overlay")
```

Acknowledgements

The authors thank Jim Shofzhal for his support regarding the Thermo Fisher RawFileReader library. We thank Sven Brehmer from Bruker Daltonics for the discussions of IMS/TOF file format. We thank Lily van de Veen for the package stickers design. We thank our colleagues at the Functional Genomics Center Zurich, the Swiss Federal Institute of Technology Zurich and the University of Zurich for their support of our work.