FT-MS Interest Group Workshop Report

"Open and Reproducible Data Analysis for FT-MS"

71st ASMS Conference on Mass Spectrometry and Allied Topics, Houston TX

June 6, 2023

Co-organizers:

David S. Butcher, National High Magnetic Field Laboratory, Tallahassee FL

Yuri E. Corilo, Pacific Northwest National Laboratory, Richland, WA

Attendance: 36 total

Description:

The 2023 FTMS interest group workshop focused on methods to use for open and reproducible data analysis of FTMS data. The session opened with a presentation by organizer David Butcher introducing open science, reproducibility, recent developments in open science, and the benefits of working reproducibly. This was followed by a presentation by Carlos Afonso (University of Rouen, Normandy, France) introducing the iC2MC project and the PyC2MC Viewer for data processing for highly complex mixture analysis. The software is intended to provide an easy-to-use open-source alternative to existing aging and closed-source solutions. Afterwards, a live demonstration of the capabilities of PyC2MC was conducted by student Maxime Sueur.

Viviana Freire-Zapata (Department of Environmental Science, University of Arizona, Tuczon, AZ) presented MetaboDirect, software designed to bridge the gap for non-programmers by providing "an open, user-friendly pipeline to analyze DI-FT-MS data". The software is intended to facilitate analysis of FT-MS data derived from highly complex samples such as soil or tissue. It is implemented in both Python and R and provides analysis scripts to users to facilitate reproducibility. Common analysis and visualization tasks are automated, and transparency and reproducibility are strongly emphasized.

Finally, James S. Prell (Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR) presented "iFAMS Quant: Fourier- and Gàbor-Transform-Based Software for Mass Spectrometry Data Analysis". The software is intended for deconvolution of "severely congested" or overlapping mass spectral signals. It can also be applied to intact protein quantitation and macromolecular mass defect analysis and has a variety of usability features intended to facilitate reproducibility, such as the export of parameters necessary for reproduction in human-readable files.

After the presentations, the presenters were brought to the front of the room for a panel discussion for the remaining time of about 45 minutes. The conversation was varied, with many audience members participating. The subject of metadata was a significant

concern, especially the concept of maintaining metadata associated with FT-MS data through varying analysis pipelines. This can be complicated further by the different formats and metadata standards employed by instrument vendors. Maintaining provenance through data processing pipelines was also seen as crucial – "cradle-to-grave" provenance data on the sample, the analysis, and the results were discussed. The opinion was expressed that collecting and maintaining this information might be best handled in a top-down fashion by an organization providing the necessary cyberinfrastructure.

The major issue of standardizing the processing of transients was raised. Processing of raw FT-MS transients tends to vary across different vendors and is generally closed-source, making it difficult to ensure reproducibility of this fundamental step of FT-MS analysis. Cross OS compatibility of FTMS analysis software was another fundamental problem which is exacerbated by operating system vendors who may break essential functionality while providing updates. Containerization and cloud computing are emerging solutions to the problems of OS compatibility.

Next year's workshop will be organized by Yuri Corilo and a new co-organizer to be decided later.