Annotated Recurrent Unidentified Spectra (ARUS)

To assist in the compound identification problem, the NIST Mass Spectrometry Data Center has developed a novel type of mass spectral library, one that includes all recurrent unidentified mass spectra in a material. Unlike traditional spectral libraries, which consist of reference spectra of known compounds derived from neat standards, these libraries are derived from recurring spectra of unknown identity in the target material itself, where spectra are extracted, clustered, and where possible annotated prior to entry into a library. Building the library itself follows a similar methodological procedure to the one described for libraries of neat compounds, though with a different set of spectrum and measurement annotation.

In general, this type of library can be useful in many usual tasks of 'omics studies (i) answering where, how often, and in what conditions certain ions are observed, (ii) assigning class ID for compounds not in current tandem mass spectral libraries or not commercially available, (iii) connecting samples in an unambiguous way for control-case studies or interlaboratory comparisons (each molecular feature is represented by a spectrum in the library).

Getting Started

To run the ARUS library, download the installation program for the NIST Search Software and the library from the NIST Website (https://chemdata.nist.gov/). A copy of the library must be present in the folder "MSSEARCH" before opening the browser.

Most datasets were generated using a Fusion Lumos Orbitrap. Eventually, data from other instruments such as Agilent and Waters QTOF was also collected. Although no systematic comparison has been made between data from different instruments, the library coverage for all instruments is similar. However, mass accuracy and ranges need to be adjusted accordingly in order to yield similar library scores for the same ions.

Prerequisites

Downloaded the NIST Search Software and libraries from the NIST Website (https://chemdata.nist.gov/).

Installing

The installation of the search program is straightforward; however, a detailed manual can be found on the website accompanying the software.

Running the tests

A human plasma data-dependent dataset can be found at https://nvlpubs.nist.gov/nistpubs/jres/121/jres.121.022.pdf and used for testing the installation and library searching.

Built With

- NIST pipeline
- NISTms and MSpepSearch

Contributions

All software and datasets were developed at the NIST Mass Spectrometry Data Center. For submitting requests or external contributions, please contact:

massspec@nist.gov

Versioning

Versions 3.0, other versions are available upon request.

Authors

NIST Mass Spectrometry Data Center.

License

NIST Licensing Policy. The library is offered without warranty and will be in continuous development in the future (http://chmdatafnx.nist.gov/dokuwiki/doku.php?id=chemdata:arus).