Package 'msdata'

April 14, 2020

Version 0.26.0

Title Various Mass Spectrometry raw data example files

Author Steffen Neumann <sneumann@ipb-halle.de>, Laurent Gatto
<laurent.gatto@uclouvain.be> with contriutions from Johannes Rainer

Maintainer
Steffen Neumann <sneumann@ipb-halle.de>, Laurent Gatto <laurent.gatto@uclouvain.be>

Depends R (>= 2.10)

Suggests xcms, mzR, MSnbase

ZipData no

Description Ion Trap positive ionization mode data in mzData file format. Subset from 500-850 m/z and 1190-1310 seconds, incl. MS2 and MS3, intensity threshold 100.000. Extracts from FTICR Apex III, m/z 400-450. Subset of UPLC - Bruker micrOTOFq data, both mzData, mzML and mz5. LC-MSMS and MRM files from proteomics experiments. PSI mzIdentML example files for various search engines.

biocViews ExperimentData, MassSpectrometryData

License GPL (>= 2)

git_url https://git.bioconductor.org/packages/msdata

git_branch RELEASE_3_10

git_last_commit 5eea689

git_last_commit_date 2019-10-29

Date/Publication 2020-04-14

R topics documented:

msdata *Sample FTICR, LC/MS and MS^n data*

Description

x object containing a subset of LC/MS raw data from a Thermo Finnigan LCQ Deca XP The data is a subset from 500-850 m/z and 1190-1310 seconds, incl. MS2 and MS3, intensity threshhold 100.000. It was collected in positive ionization mode.

xs object containing a subset of FTICR data from a Bruker APex III FTICR. The data is a subset from 400-450 m/z, collected in positive ionization mode.

Usage

```
data(xs)
```

Format

The format is:

```
xs
```

Details

The corresponding raw mzdata files are located in the `fticr` and `iontrap` subdirectory of this package.

See Also

xcmsSet, xcmsRaw

Examples

```
## The directory with the mzData LC/MS files
data(xs)
mzdatapath <- file.path(find.package("msdata"), "iontrap")
mzdatapath
files <- list.files(mzdatapath, recursive = TRUE, full.names = TRUE)
files

if (require(xcms)) {

## xcmsSet Summary
show(xs)

## Access raw data file
x <- xcmsRaw(files[1])
x

}
```

proteomics *Proteomics data in* msdata

Description

This function returns proteomics mass spectrometry files. These files are all stored in the `proteomics` directory in the `msdata` package. Each file/data is described in more details below.

Usage

```
proteomics(...)
```

Arguments

...	Additional arguments passed to `list.files`.

Details

- `TMT_Erwinia_1uLSike_Top10HCD_isol2_45stepped_60min_01.mzML.gz`: A LC-MSMS data file containing iTRAQ 4-plex data. The data is described in more details in Gatto L. and Christoforou A. *Using R and Bioconductor for proteomics data analysis* (PMID 23692960). This file only contains a subset of the fill data (spectra 1002 to 1510) and was generated from the full data using `msconvert` (ProteoWizard release: 3.0.9283 (2016-1-11)) using following command

  ```
  msconvert TMT_Erwinia_1uLSike_Top10HCD_isol2_45stepped_60min_01-20141210.mzML
  --filter "index [1002,1510]" -o subset
  ```

  The complete file is `TMT_Erwinia_1uLSike_Top10HCD_isol2_45stepped_60min_01-20141210.mzML.gz`, also available here, and can also be downloaded from the ProteomeXchange `PXD000001` project (see the `rpx` package).

  An MS2 identification file, `ident/TMT_Erwinia_1uLSike_Top10HCD_isol2_45stepped_60min_01-20141210.m` generated searching the raw data against the *Erwinia carotovora* database (see reference above) is also available through the `ident` function.

- `MS3TMT10_01022016_32917-33481.mzML.gz`:A subset of 565 spectra from a currenly unpublished TMT 10-plex experiment run on an Thermo Orbitrap Lumos with synchronous precursor selection (SPS) MS3. Only the MS2 spectra were centroided during convertion using `msconvert` (ProteoWizard release: 3.0.9283 (2016-1-11)) using vendor libraries.

- `MS3TMT11.mzML`:A subset of 994 spectra from a currenly unpublished MS3 SPS TMT 11-plex experiment converted to mzML using `msconvert`. The file contains 30, 482 and 482 MS1, MS2 and MS3 spectra, respectively. The MS1 spectra are in profile mode; other MS levels are centroided. See *Sensitive and Accurate Quantitation of Phosphopeptides Using TMT Isobaric Labeling Technique* for details about the acquisition method.

  An feature data containing identification data is available with `data(fdms3tmt11)`, which can be used to directly update the feature data, as shown in the example below.

- MRM-standmix-5.mzML.gz:Sample from mouse brain acquired by HILIC ESI-QqQ/MS in Dynamic multiple reaction monitoring mode (MRM). HPLC system was a 1290 Infinity (Agilent Technologies) coupled to ion-Funnel Triple quadrupole 6490 mass spectrometer (Agilent Technologies). This file was contributed by Xavi Domingo-Almenara from theThe Scripps Research Institute, San Diego, CA.

Value

A `character` with file names.

Author(s)

Laurent Gatto <laurent.gatto@uclouvain.be>

See Also

For more access to mass spectrometry-based proteomics data, see the `rpx` and `ProteomicsAnnotationHubData` packages.

Examples

```
## raw data files
(f <- proteomics(full.names = TRUE))

library("mzR")
openMSfile(f[2])

library("MSnbase")
## The MS3 TMT11 raw data
(fms3 <- proteomics(full.names = TRUE, pattern = "MS3TMT11.mzML"))
ms3 <- readMSData(fms3, mode = "onDisk")
ms3

## Additional feature metadata
data(fdms3tmt11)
names(fdms3tmt11)

fData(ms3) <- fdms3tmt11
validObject(ms3)

## identification data file
ident(full.names = TRUE)

## quantiative data files
quant(full.names = TRUE)
```

`sciexdata` *AB Sciex LC-MS data files*

Description

The mzML files in the `sciex` directory in the `msdata` package represent profile-mode LC-MS data of pooled human serum samples (the same pool being measured). The samples were analyzed by ultra high-performance liquid chromatography (UHPLC; Agilent 1290) coupled to a Q-TOF mass spectrometer (TripleTOF 5600+ AB Sciex). The chromatographic separation was based in hydrophilic interaction liquid chromatography (HILIC) and performed using an Waters Acquity BEH Amide, 100 x 2.1 mm column.

The mass specctrometer was operated in full scan mode in the mass range from 50 to 1000 m/z and with an accumulation time of 250 ms. The files represent a subset of spectra/scans from m/z 105 to 134 and from retention time 0 to 260 seconds. The files were generated in the same LC-MS run, but from different injections. Details on the individual files are provided below.

Details

- 20171016_POOL_POS_1_105-134.mzML profile-mode LC-MS data of pooled human serum samples. Injection index: 1.
- 20171016_POOL_POS_3_105-134.mzML profile-mode LC-MS data of pooled human serum samples. Injection index: 19.

Author(s)

Sigurdur Smarason, Giuseppe Paglia and Johannes Rainer

Examples

```
## List the files in the sciex folder
dir(system.file("sciex", package = "msdata"))
```

Index