

# Package ‘msdata’

April 10, 2025

**Version** 0.47.0

**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 mzML 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 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)

**Encoding** UTF-8

**git\_url** <https://git.bioconductor.org/packages/msdata>

**git\_branch** devel

**git\_last\_commit** bb50ac0

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.21

**Date/Publication** 2025-04-10

## Contents

CE-MS . . . . .	2
msdata . . . . .	2
proteomics . . . . .	3
sciexdata . . . . .	5

---

CE-MS

*CE-MS test data*

---

### Description

The CE-MS test files consist of two files, i.e. "CEMS\_10ppm.mzML" and "CEMS\_25ppm.mzML". The data contains CE-MS runs of a standard mixture that contains e.g. Lysine (at 10 ppm and 25 ppm, respectively) and the neutral EOF marker Paracetamol (50 ppm). The data was acquired on a 7100 capillary electrophoresis system from Agilent Technologies, coupled to an Agilent 6560 IM-QToF-MS. CE Separation was performed using a 80 cm fused silica capillary with an internal diameter of 50  $\mu\text{m}$  and external diameter of 365  $\mu\text{m}$ . The Background Electrolyte was 10 % acetic acid and separation was performed at +30 kV and a constant pressure of 50 mbar. MS detection was performed in positive ionization mode.

The raw data were then converted to an open-source ".mzML" format (Proteowizzard) and load into R via the `MSnBase::readMSData()` function. In order to reduce data size, the test data was subsequently cutted in migration time and m/z range using `filterRt(rt = c(400, 900))` and `filterMz(mz = c(147.1, 152.0))` from `MSnBase`

### Author(s)

Liesa Salzer

---

msdata

*Sample FTICR, LC/MS and MS<sup>n</sup> 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 threshold 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 mzML files are located in the `fticr-mzML` and `iontrap` subdirectory of this package.

## See Also

[xcmsSet](#), [xcmsRaw](#)

## Examples

```
## The directory with the mzML LC/MS files
data(xs)
mzMLpath <- file.path(find.package("msdata"), "iontrap")
mzMLpath
files <- list.files(mzMLpath, 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 TMT6 6-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 full 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.mzid`, 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 currently unpublished TMT 10-plex experiment run on an Thermo Orbitrap Lumos with synchronous precursor selection (SPS) MS3. Only the MS2 spectra were centroided during conversion using `msconvert` (ProteoWizard release: 3.0.9283 (2016-1-11)) using vendor libraries.
- MS3TMT11.mzML: A subset of 994 spectra from a currently 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 the 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)

## quantitative 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 spectrometer 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

## \* datasets

msdata, [2](#)

sciexdata, [5](#)

CE-MS, [2](#)

fdms3tmt11 (proteomics), [3](#)

ident (proteomics), [3](#)

list.files, [3](#)

msdata, [2](#)

proteomics, [3](#)

quant (proteomics), [3](#)

sciexdata, [5](#)

x (msdata), [2](#)

xcmsRaw, [3](#)

xcmsSet, [3](#)

xs (msdata), [2](#)