

# Package ‘topdownrdata’

October 10, 2024

**Version** 1.27.0

**Title** Example Files for the topdownr R Package

**Description** Example data for the topdownr package generated on a Thermo Orbitrap Fusion Lumos MS device.

**Depends** topdownr

**biocViews** ExperimentData, MassSpectrometryData

**License** GPL (>= 3)

**NeedsCompilation** no

**URL** <https://github.com/sgibb/topdownrdata/>

**BugReports** <https://github.com/sgibb/topdownrdata/issues/>

**RoxygenNote** 6.0.1.9000

**Roxygen** list(markdown=TRUE)

**Encoding** UTF-8

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

**git\_branch** devel

**git\_last\_commit** 36f7489

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.20

**Date/Publication** 2024-10-10

**Author** Sebastian Gibb [aut, cre],  
Pavel Shliaha [aut],  
Ole Nørregaard Jensen [aut]

**Maintainer** Sebastian Gibb <[mail@sebastiangibb.de](mailto:mail@sebastiangibb.de)>

## Contents

topdownrdata-package . . . . .	2
topDownDataPath . . . . .	4
<b>Index</b>	<b>6</b>

---

topdownrdata-package *Example Data for the topdownr package.*

---

## Description

This package contains example files accompanying the topdownr.

## Details

It has just one function `topDownDataPath()` that returns the file path to the 5 example protein datasets.

Each dataset has four different categories of files:

- One `.fasta` file containing the protein sequence.
- Multiple `.experiments.csv`, `.txt`, and `.mzML` files (the same number of files for each of the three types):
  - The `.experiments.csv` files contain the information about the used method and the settings of the mass spectrometer (fragmentation conditions).
  - The `.txt` scan header files contain (additional) information about the spectra (monoisotopic  $m/z$ , ion injection time, ...).
  - The `.mzML` files contain the deconvoluted spectra.

In total this package has 341 files: a `.fasta` file for each protein (5) and 20 files of each of the three method/spectra information files for every protein except for the *bovine carbonic anhydrase* and *C3a recombinant protein* which have 26 of each.

The topdownr package needs all the four file types. The sequence information of the `.fasta` file is used to calculate the fragmentation *in-silico*. The theoretical fragments are matched against the experimental seen fragments that are stored in the `.mzML` files. In the next step the fragmentation data have to be combined with the general information about spectra and the fragmentation condition from the `.txt` scan header and the `.experiments.csv` method files, respectively.

In combination these information could be used to investigate fragmentation conditions and to find the one (or more) that maximise the overall fragment coverage. Please see a small example on the end of this manual page and a full featured example analysis in the topdownr analysis vignette: `vignette("analysis", package="topdownr")`.

The `.meth` files were created with the following command:

```
library("topdownr")

writeMethodXmls(defaultMs1Settings(LastMass=1600),
                defaultMs2Settings(),
                ## mass/z adapted to protein of interest (see table)
                ## z is currently not supported by the Thermo software,
                ## setting to 1.
                mz=cbind(mass=c(745.2, 908.0, 1162.0), z=c(1, 1, 1)),
                groupBy=c("replication", "ETDReactionTime"),
                replications=2,
                pattern="method_CA3_\\%s.xml")
```

**General Information:**

protein name	uniprot accession	product number	modifications	monoisotopic mass
horse myoglobin	P68082	sigma M1882	Met-loss	
bovine carbonic anhydrase	P00921	sigma C2522	Met-loss + Acetyl	
histone H3.3	P84243	NEB M2507S	Met-loss	
histone H4	P62805	NEB M2504S	Met-loss	
C3a recombinant protein	P01024 part (672-748)	recombinantly expressed	carbamidomethyl	

All 5 proteins were infused into a Thermo Orbitrap Fusion Lumos at 600 nl/minute in 50 % acetonitrile 0.1 FS360-20-10-5-6.35CT emitter.

**M/Z used:**

protein name	m/z 1	m/z 2	m/z 3
horse myoglobin	707.3/24	893.1/19	1211.7/14
bovine carbonic anhydrase	745.2/39	908.0/32	1162.0/25
histone H3.3	563.8/27	691.8/22	894.9/17
histone H4	562.7/20	703.2/16	937.3/12
C3a recombinant protein	745.2/17	908.0/14	1162.0/11

**Author(s)**

Pavel Shliaha <pavels@bmb.sdu.dk>, Sebastian Gibb <mail@sebastiangibb.de>

**References**

<https://github.com/sgibb/topdownrdata/>

**See Also**

[topDownDataPath\(\)](#), [topdownr-package](#),  
 Vignettes for the generation `vignette("data-generation", package="topdownr")` and analysis  
 of these data `vignette("analysis", package="topdownr")`.  
 Website: <https://sgibb.github.io/topdownr/>

**Examples**

```
# List file categories
list.files(topdownrdata::topDownDataPath("myoglobin"))

# List all needed files
list.files(topdownrdata::topDownDataPath("myoglobin"), recursive=TRUE)

# Read files, predict fragments and combine spectra information
tds <- readTopDownFiles(
  path=topDownDataPath("myoglobin"),
  ## Use an artificial pattern to load just the fasta
  ## file and files from m/z == 1211, ETD reagent
  ## target 1e6 and first replicate to keep runtime
  ## of the example short
  pattern=".*fasta.gz$|1211_.*1e6_1"
)
```

```
# Show TopDownSet object
tds

# Filter all intensities that don't have at least 10 % of the highest
# intensity per fragment.
tds <- filterIntensity(tds, threshold=0.1)

# Filter all conditions with a CV above 30 % (across technical replicates)
tds <- filterCv(tds, threshold=30)

# Filter all conditions with a large deviation in injection time
tds <- filterInjectionTime(tds, maxDeviation=log2(3), keepTopN=2)

# Filter all conditions where fragments don't replicate
tds <- filterNonReplicatedFragments(tds)

# Normalise by TIC
tds <- normalize(tds)

# Aggregate technical replicates
tds <- aggregate(tds)

# Coerce to NCBSets (N-/C-terminal/Bidirectional) and plot fragment coverage
fragmentationMap(as(tds, "NCBSets"))
```

---

topDownDataPath

*TopDown Proteomic Datasets*

---

## Description

This function returns the path to the example files accompanying the topdownr.

## Usage

```
topDownDataPath(protein = c("myoglobin", "ca", "h3_3", "h4", "c3a"))
```

## Arguments

protein            character, name of the dataset.

## Details

See [topdownrdata-package](#) for a description of the datasets.

## Value

character, path to the directory containing the example files.

## Author(s)

Sebastian Gibb <[mail@sebastiangibb.de](mailto:mail@sebastiangibb.de)>

**See Also**

<https://sgibb.github.io/topdownr/>

**Examples**

```
topDownDataPath("myoglobin")
```

# Index

## \* **package**

topdownrdata-package, [2](#)

topDownDataPath, [4](#)

topDownDataPath(), [2](#), [3](#)

topdownr-package, [3](#)

topdownrdata (topdownrdata-package), [2](#)

topdownrdata-package, [2](#), [4](#)