Package 'sfi'

April 15, 2026

Type Package

Title Data analysis for Single File Injections (SFIs) mode LC-MS analysis

Version 0.99.5

Description Data analysis for Single File Injections(SFIs) mode LC-MS analysis. In SFIs mode, pooled samples are initially injected to serve as reference peaks for subsequent analyses. Repeated injections of individual samples are then performed at fixed time intervals using isocratic elution. This package provides the functions to analyze data from SFIs mode including peak picking and peak reassignment.

URL https://github.com/yufree/sfi

BugReports https://github.com/yufree/sfi/issues/new

License MIT + file LICENSE

Depends R (>= 4.5.0)

LinkingTo Rcpp

Imports Rcpp, enviGCMS, stats, mzR, rmarkdown, methods, SummarizedExperiment, S4Vectors

Suggests knitr, data.table, BiocStyle, ggplot2, MsCoreUtils, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

biocViews MassSpectrometry, Metabolomics, FeatureExtraction

Encoding UTF-8

RoxygenNote 7.3.3

LazyData false

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

git_branch devel

git_last_commit e0897f4

git_last_commit_date 2026-03-18

Repository Bioconductor 3.23

Date/Publication 2026-04-14

Author Miao YU [aut, cre] (ORCID: <https://orcid.org/0000-0002-2804-6014>)

Maintainer Miao YU <yufreecas@gmail.com>

Contents

`find_2d_peaks` *Feature extraction core function*

Description

This function finds local max peaks on m/z-retention time 2D plane.

Usage

```
find_2d_peaks(
  mz,
  rt,
  intensity,
  ppm = 5,
  deltart = 5,
  snr = 3,
  mz_bins = NULL,
  rt_bins = NULL
)
```

Arguments

`mz`	Numeric vector of m/z values.
`rt`	Numeric vector of retention times.
`intensity`	Numeric vector of intensities corresponding to m/z and rt values.
`ppm`	Numeric. Parts per million tolerance for m/z matching. Default is 5.

deltart	Numeric. Tolerance for retention time matching. Default is 5.
snr	Numeric. signal to ratio to find peaks.
mz_bins	Numeric. m/z bins. Default 50000.
rt_bins	Numeric. retention time bins. Default 100.

Value

A data frame containing m/z, retention time, and intensity of identified peaks.

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
```

---

find_peaks_low_res *Find peaks in low-resolution data using the 2D peak finding algorithm*

Description

This function adapts the fast ‘find_2d_peaks‘ function for use with low-resolution (unit mass) data. It does this by first aggregating the signal at each integer mass and then calling the 2D peak finder.

Usage

```
find_peaks_low_res(mz, rt, intensity, deltart = 5, snr = 3)
```

Arguments

mz	A numeric vector of mass-to-charge ratios.
rt	A numeric vector of retention times.
intensity	A numeric vector of intensities.
deltart	Numeric. Tolerance for retention time matching. Default is 5.
snr	Numeric. Signal-to-noise ratio to find peaks. Default is 3.0.

Value

A data frame with columns ‘mz‘, ‘rt‘, and ‘intensity‘, representing the detected peaks. The ‘mz‘ values are integer masses.

Examples

```
data(sfi)
peaks <- find_peaks_low_res(
    mz = sfi$mz, rt = sfi$rt,
    intensity = sfi$intensity
)
```

`getidelta` *Optimize Delta Retention Time*

Description

This function optimizes the delta retention time (idelta) using a binary search approach.

Usage

```
getidelta(mz, rt, ...)

## S3 method for class 'sfi_peaks'
getidelta(mz, rt = NULL, ...)

## Default S3 method:
getidelta(
  mz,
  rt,
  qcmz,
  qcrt,
  idelta = 60,
  shift = 0,
  ppm = 5,
  deltart = 5,
  window = 600,
  n = 160,
  tol = 0.03,
  max_iter = 100,
  ...
)
```

Arguments

`mz`	Numeric vector of m/z values or an object of class ‘sfi_peaks‘.
`rt`	Numeric vector of retention times.
`...`	Additional arguments passed to methods.
`qcmz`	Numeric vector of QC m/z values.
`qcrt`	Numeric vector of QC retention times.
`idelta`	Initial delta retention time guess. Default is 60.
`shift`	Numeric. Shift applied to idelta. Default is 0.
`ppm`	Numeric. Parts per million tolerance for m/z matching. Default is 5.
`deltart`	Numeric. Tolerance for retention time matching. Default is 5.
`window`	Numeric. Retention time window. Default is 600.
`n`	Integer. Number of iterations or samples. Default is 160.
`tol`	Numeric. Tolerance for binary search convergence. Default is 0.03.
`max_iter`	Integer. Maximum number of binary search iterations. Default is 100.

Value

Optimized delta retention time (idelta).

Methods (by class)

- `getidelta(sfi_peaks)`: Method for sfi_peaks object
- `getidelta(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
delta_opt <- getidelta(peak$mz, peak$rt,qcmz=195.0876,qcrt=74,window=632,idelta=90)
```

---

getmzml *Read mzML File and Extract m/z, Retention Time, and Intensity*

Description

Read mzML File and Extract m/z, Retention Time, and Intensity

Usage

```
getmzml(path)
```

Arguments

`path`	path of SFI mzML file.

Value

A data frame containing m/z, retention time and intensity.

Examples

```
# Load demo data
data(sfi)
head(sfi)
# In practice, you would use a real mzML file path:
# peak <- getmzml("path/to/your/file.mzML")
# The function returns a data frame with m/z, retention time, and intensity columns
```

getsff *Cluster and Pair m/z and Retention Time Features*

Description

This function clusters m/z values based on Manhattan distance and pairs features within clusters.

Usage

```
getsff(mz, rt, ...)

## S3 method for class 'sfi_peaks'
getsff(mz, rt = NULL, ...)

## Default S3 method:
getsff(mz, rt, ppm = 5, minn = 2, refmz = NULL, ...)
```

Arguments

mz	Numeric vector of m/z values or an object of class ‘sfi_peaks‘.
rt	Numeric vector of retention times corresponding to m/z values.
...	Additional arguments passed to methods.
ppm	Numeric. Parts per million tolerance for m/z matching. Default is 5.
minn	Integer. Minimum number of features in a cluster to be retained. Default is 2.
refmz	Optional numeric vector of reference m/z values for alignment. Default is NULL.

Value

A data frame containing paired m/z and retention time values with their differences:

- mz1: m/z of the first feature in the pair.
- rt1: retention time of the first feature in the pair.
- mz2: m/z of the second feature in the pair.
- rt2: retention time of the second feature in the pair.
- pmr: absolute difference in retention time (Pair Mass Retention).
- pmd: absolute difference in m/z (Pair Mass Difference).

Methods (by class)

- `getsff(sfi_peaks)`: Method for sfi_peaks object
- `getsff(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
sff_features <- getsff(peak$mz, peak$rt)
```

getsfm *Generate Sample Feature Matrix (SFM)*

Description

This function generates a Sample Feature Matrix (SFM) by aligning and filtering sample peaks against QC peaks. The SFM contains features extracted from individual samples within the single file injection.

Usage

```
getsfm(mz, rt, intensity, ...)

## S3 method for class 'sfi_peaks'
getsfm(mz, rt = NULL, intensity = NULL, ...)

## Default S3 method:
getsfm(
  mz,
  rt,
  intensity,
  idelta = 60,
  windows = 600,
  qcseq = c(1, 1, 0, 1, 1, 0, 1, 1, 0),
  deltart = 5,
  ppm = 5,
  minn = 1,
  n = 160,
  ...
)
```

Arguments

mz	Numeric vector of m/z values or an object of class ‘sfi_peaks‘.
rt	Numeric vector of retention times.
intensity	Numeric vector of intensities corresponding to m/z and rt values.
...	Additional arguments passed to methods.
idelta	Numeric. Initial delta retention time. Default is 60.
windows	Numeric. Retention time window. Default is 600.
qcseq	Integer vector indicating QC samples. Default is c(1, 1, 0, 1, 1, 0, 1, 1, 0).
deltart	Numeric. Tolerance for retention time matching. Default is 5.
ppm	Numeric. Parts per million tolerance for m/z matching. Default is 5.
minn	Integer. Minimum number of QC samples required. Default is 1.
n	Integer. Number of samples for delta optimization. Default is 160.

Value

A data frame containing the aligned and filtered sample features with the following columns:

- mz: m/z of the feature in the sample.
- rt: retention time of the feature in the sample (global).
- srt: relative retention time of the feature within the sample injection window.
- sampleidx: index of the sample injection.
- intensity: intensity of the feature.
- qcmz: m/z of the matching reference QC feature.
- qcrt: retention time of the matching reference QC feature.
- shiftrt: absolute difference between sample srt and QC reference retention time.
- ppmshift: absolute difference in ppm between sample m/z and QC reference m/z.

The row names of the data frame are set to the sample index (injection number).

Methods (by class)

- `getsfm(sfi_peaks)`: Method for sfi_peaks object
- `getsfm(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
sfm_df <- getsfm(peak$mz, peak$rt, peak$intensity,idelta=92,windows=632,minn=6,n=158,deltart=10)
```

`getwindow` *Determine Optimal Retention Time Window*

Description

This function calculates the optimal retention time window based on QC sequences and m/z/rt data.

Usage

```
getwindow(mz, rt, ...)

## S3 method for class 'sfi_peaks'
getwindow(mz, rt = NULL, ...)

## Default S3 method:
getwindow(
  mz,
  rt,
  lower = 620,
```

```
  upper = 650,
  ppm = 5,
  minn = 1,
  qcseq = c(1, 1, 0, 1, 1, 0, 1, 1, 0),
  ...
)
```

Arguments

mz	Numeric vector of m/z values or an object of class 'sfi_peaks'.
rt	Numeric vector of retention times.
...	Additional arguments passed to methods.
lower	Numeric. Lower bound for the retention time window. Default is 620.
upper	Numeric. Upper bound for the retention time window. Default is 650.
ppm	Numeric. Parts per million tolerance for m/z matching. Default is 5.
minn	Integer. Minimum number of features in a QC cluster. Default is 1.
qcseq	Integer vector. QC sequence indicating which samples are QC. Default is c(1, 1, 0, 1, 1, 0, 1, 1, 0).

Value

Numeric value representing the optimal retention time window.

Methods (by class)

- `getwindow(sfi_peaks)`: Method for sfi_peaks object
- `getwindow(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
window_opt <- getwindow(peak$mz, peak$rt)
```

---

get_qc_features *Generate Quality Control Feature List*

Description

This function generates a list of features found in Quality Control (QC) samples by aligning QC and matrix samples and filtering based on detection frequency criteria.

Usage

```
get_qc_features(mz, rt, intensity, ...)

## S3 method for class 'sfi_peaks'
get_qc_features(mz, rt = NULL, intensity = NULL, ...)

## Default S3 method:
get_qc_features(
  mz,
  rt,
  intensity,
  idelta = 60,
  windows = 600,
  qcseq = c(1, 1, 0, 1, 1, 0, 1, 1, 0),
  deltart = 5,
  ppm = 5,
  minn = 6,
  ...
)
```

Arguments

mz	Numeric vector of m/z values or an object of class ‘sfi_peaks‘.
rt	Numeric vector of retention times.
intensity	Numeric vector of intensities corresponding to m/z and rt values.
...	Additional arguments passed to methods.
idelta	Numeric. Initial delta retention time. Default is 60.
windows	Numeric. Retention time window. Default is 600.
qcseq	Integer vector indicating QC samples. Default is c(1, 1, 0, 1, 1, 0, 1, 1, 0).
deltart	Numeric. Tolerance for retention time matching. Default is 5.
ppm	Numeric. Parts per million tolerance for m/z matching. Default is 5.
minn	Integer. Minimum number of QC samples required. Default is 6.

Value

A data frame containing filtered QC features with the following columns:

- mzqc: aligned m/z of the QC feature.
- rtqc: aligned retention time of the QC feature.
- intensity: intensity of the feature in the specific QC sample.
- sampleidx: index of the QC sample injection.
- idxq: unique identifier for the QC feature group (mz rt).

The row names of the data frame are set to the sample index (injection number), with suffixes to ensure uniqueness.

Methods (by class)

- `get_qc_features(sfi_peaks)`: Method for sfi_peaks object
- `get_qc_features(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
qc_features <- get_qc_features(peak$mz, peak$rt, peak$intensity,
                               idelta=92.25,windows=632.11,minn=6,deltart=10)
```

`get_sfi_params` *Quality Control for Mass Spectrometry Data*

Description

This function performs quality control (QC) on mass spectrometry data by aligning QC and sample features.

Usage

```
get_sfi_params(mz, rt, intensity, ...)

## S3 method for class 'sfi_peaks'
get_sfi_params(mz, rt = NULL, intensity = NULL, ...)

## Default S3 method:
get_sfi_params(
  mz,
  rt,
  intensity,
  idelta = 60,
  window = 600,
  qcseq = c(1, 1, 0, 1, 1, 0, 1, 1, 0),
  deltart = 5,
  ppm = 5,
  minn = 1,
  n = 160,
  tol = 0.03,
  max_iter = 100,
  wlower = 620,
  wupper = 650,
  ...
)
```

Arguments

mz	Numeric vector of m/z values or an object of class ‘sfi_peaks‘.
rt	Numeric vector of retention times.
intensity	Numeric vector of intensities corresponding to m/z and rt values.
...	Additional arguments passed to methods.
idelta	Numeric. Initial delta retention time. Default is 60.
window	Numeric. Retention time window. Default is 600.
qcseq	Integer vector indicating QC samples. Default is c(1, 1, 0, 1, 1, 0, 1, 1, 0).
deltart	Numeric. Tolerance for retention time matching. Default is 5.
ppm	Numeric. Parts per million tolerance for m/z matching. Default is 5.
minn	Integer. Minimum number of QC samples required. Default is 1.
n	Integer. Number of samples for delta optimization. Default is 160.
tol	Numeric. Tolerance for binary search in delta optimization. Default is 0.03.
max_iter	Integer. Maximum iterations for binary search. Default is 100.
wlower	Numeric. Lower bound for window determination. Default is 620.
wupper	Numeric. Upper bound for window determination. Default is 650.

Value

A named numeric vector containing the optimal window and delta retention time:

- window: The optimized retention time window.
- idelta: The optimized delta retention time.

Methods (by class)

- `get_sfi_params(sfi_peaks)`: Method for sfi_peaks object
- `get_sfi_params(default)`: Default method for vectors

Examples

```
data(sfi)
peak <- find_2d_peaks(mz=sfi$mz,rt=sfi$rt,intensity=sfi$intensity)
sfi_params <- get_sfi_params(peak$mz, peak$rt, peak$intensity, deltart=10)
```

run_app *Run sfi Shiny App*

Description

A function to run the shiny app for sfi package

Usage

```
run_app()
```

Value

A shiny app

sfi *Demo sfi data*

Description

Demo sfi data

Usage

```
data(sfi)
```

Format

A data.frame object with mass to charge ratio, intensity and retention time from sfi mode.

Index