# Package: pmd (via r-universe)

October 9, 2026

**Type** Package

**Title** Paired Mass Distance Analysis for GC/LC-MS Based Non-Targeted Analysis and Reactomics Analysis

**Version** 0.3.0

**Description** Paired mass distance (PMD) analysis proposed in Yu, Olkowicz and Pawliszyn (2018) <doi:10.1016/j.aca.2018.10.062> and PMD based reactomics analysis proposed in Yu and Petrick (2020) <doi:10.1038/s42004-020-00403-z> for gas/liquid chromatography–mass spectrometry (GC/LC-MS) based non-targeted analysis. PMD analysis including GlobalStd algorithm and structure/reaction directed analysis. GlobalStd algorithm could found independent peaks in m/z-retention time profiles based on retention time hierarchical cluster analysis and frequency analysis of paired mass distances within retention time groups. Structure directed analysis could be used to find potential relationship among those independent peaks in different retention time groups based on frequency of paired mass distances. Reactomics analysis could also be performed to build PMD network, assign sources and make biomarker reaction discovery. GUIs for PMD analysis is also included as 'shiny' applications.

**URL** https://yufree.github.io/pmd/

**BugReports** https://github.com/yufree/pmd/issues

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**Suggests** knitr, shiny, rmarkdown

**VignetteBuilder** knitr

**Depends** R (>= 3.5.0)

**Imports** RColorBrewer, stats, graphics, utils, igraph, enviGCMS, data.table

**Config/roxygen2/version** 8.0.0

**Config/pak/sysreqs** libglpk-dev libxml2-dev

**Repository** https://ymgroup.r-universe.dev

**Date/Publication** 2026-10-07 03:31:20 UTC

**RemoteUrl** https://github.com/yufree/pmd

**RemoteRef** HEAD

**RemoteSha** 5a29a96d12fce9e9f05e26ae8f1b5e4357012924

# Contents

getcda . . . 3
getchain . . . 4
getchainseq . . . 5
getcorpseudospectrum . . . 7
gethomolog . . . 8
getms2pmd . . . 9
getmspmd . . . 9
getnse . . . 10
getpaired . . . 10
getpmd . . . 11
getpmddf . . . 12
getposneg . . . 13
getpseudospectrum . . . 13
getrda . . . 14
getreact . . . 15
getsda . . . 16
getsim . . . 17
getstd . . . 18
gettarget . . . 19
globalstd . . . 19
hmdb . . . 20
keggrall . . . 21
MaConDa . . . 21
omics . . . 22
parse_pmd_pattern . . . 22
pcasf . . . 23
plotcn . . . 24
plotpaired . . . 24
plotrtg . . . 25
plotsda . . . 26
plotstd . . . 26
plotstdrt . . . 27
plotstdsda . . . 28
pmdchain . . . 28
runPMD . . . 29
runPMDnet . . . 29
sda . . . 30

spmeinvivo . . . . . 30

**Index** **31**

---

`getcda` *Perform correlation directed analysis for peaks list.*

---

## Description

Perform correlation directed analysis for peaks list.

## Usage

```
getcda(list, corcutoff = 0.9, rtcutoff = 10, accuracy = 4)
```

## Arguments

| | |
|---|---|
| `list` | a list with mzrt profile |
| `corcutoff` | cutoff of the correlation coefficient, default NULL |
| `rtcutoff` | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| `accuracy` | measured mass or mass to charge ratio in digits, default 4 |

## Value

list with correlation directed analysis results

## See Also

`getsda`,`getrda`

## Examples

```
data(spmeinvivo)
cluster <- getpseudospectrum(spmeinvivo)
cbp <- enviGCMS::getfilter(cluster,rowindex = cluster$stdmassindex2)
cda <- getcda(cbp)
```

getchain *Get reaction chain for specific mass to charge ratio*

## Description

Get reaction chain for specific mass to charge ratio

## Usage

```
getchain(
  list,
  diff,
  mass,
  digits = 2,
  accuracy = 4,
  rtcutoff = 10,
  corcutoff = 0.6,
  ppm = 25
)
```

## Arguments

| | |
|---|---|
| list | a list with mzrt profile |
| diff | paired mass distance(s) of interests |
| mass | a specific mass for known compound or a vector of masses. You could also input formula for certain compounds |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |
| rtcutoff | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| corcutoff | cutoff of the correlation coefficient, default 0.6 |
| ppm | all the peaks within this mass accuracy as seed mass or formula |

## Value

a list with mzrt profile and reaction chain dataframe

## Examples

```
data(spmeinvivo)
# check metabolites of C18H39NO
pmd <- getchain(spmeinvivo,diff = c(2.02,14.02,15.99),mass = 286.3101)
# remove the retention time for mass only data
spmeinvivo$rt <- NULL
pmd <- getchain(spmeinvivo,diff = c(2.02,14.02,15.99),mass = 286.3101)
```

`getchainseq` *Get reaction pathway chains matching an ordered PMD pattern with quantifiers*

## Description

Searches a feature network for directed paths whose successive mass differences follow a user-specified PMD pattern. Unlike `getchain`, which extracts the connected component reachable by any PMD in a set, this function matches an *ordered* sequence of PMDs, with support for wildcards and regex-style quantifiers. This unifies three use cases: fixed reaction sequences (e.g. glycosylation followed by dehydration), homologous series (repeated `+CH2`, PEG units, etc.), and paths with unknown intermediates (wildcard steps).

## Usage

```
getchainseq(
  list,
  pattern = NULL,
  db = NULL,
  mass = NULL,
  digits = 4,
  rtcutoff = 10,
  corcutoff = 0.6,
  ppm = 25,
  rtdir = "any",
  max_paths = 1e+05,
  allow_cycles = FALSE
)
```

## Arguments

| | |
|---|---|
| `list` | a pmd-style list with `mz`, `rt` (optional), `data` |
| `pattern` | numeric vector, list of step specs, or DSL string (see `parse_pmd_pattern`); omit when `db` is supplied |
| `db` | optional chain database (e.g. `data(pmdchain)`) for known-pathway screening: instead of one pattern, every named chain in `db` is run as an ordered query and the matches are summarized with each chain's `keggcount` specificity (lower = rarer = more specific). A match is a relational annotation (co-varying, chromatographically resolved features related by the chain's mass differences), not compound identification or proof a pathway occurred. Default `NULL` |
| `mass` | optional seed mass(es) or formula(s); only paths starting within `ppm` of a seed are returned |
| `digits` | PMD matching precision, default 4 |
| `rtcutoff` | RT hierarchical-clustering cutoff for isomer grouping, default 10 |
| `corcutoff` | correlation cutoff between linked features, default 0.6; pass `NULL` to disable |

| | |
|---|---|
| ppm | seed mass ppm tolerance, default 25 |
| rtdir | RT direction per edge: "any", "increasing", "decreasing", or a numeric vector (one entry per unrolled edge, values in {-1, 0, 1}) |
| max_paths | safety cap on returned paths, default 1e5 |
| allow_cycles | if FALSE (default), paths may not revisit a node |

## Details

Pattern grammar (any of):

- Numeric vector: each element is one fixed-PMD step. Example: c(162.0528, -18.0106) = glycosylation followed by one dehydration.
- List of step specs: each step is a list with pmd (numeric, or NA for wildcard), min (default 1), max (default = min; Inf allowed).
- Character string parsed by parse_pmd_pattern.

Signs of PMDs are significant: +162 = mass gain, -162 = mass loss. This separates e.g. glycosylation from deglycosylation.

## Value

the input list with two new elements: sdacseq (a data.table of matched paths) and pattern (the normalized pattern). sdacseq columns: n1..nK (node indices; NA-padded for shorter paths in variable-length matches), path_len, mz_1..mz_K, rt_1..rt_K (if RT present), pmd_1..pmd_{K-1} (observed PMDs), and cor_mean. In database mode (db supplied) it instead returns chainsearch (a data frame, one row per database chain: chain_id, name, class, nstep, pattern, specificity, n_found, example, ordered by n_found) and chainmatches (per-chain matched-path tables for chains with hits).

## See Also

getchain, gethomolog, parse_pmd_pattern

## Examples

```
## Not run:
data(spmeinvivo)

# Fixed sequence: glycosylation then dehydration
r1 <- getchainseq(spmeinvivo, c(162.0528, -18.0106))

# Homologous series: 2+ CH2 extensions
r2 <- getchainseq(spmeinvivo,
                  list(list(pmd = 14.0157, min = 2, max = Inf)))

# DSL string with wildcards and quantifiers
r3 <- getchainseq(spmeinvivo,
                  "162.0528; -18.0106{0,3}; 14.0157+")

# Anchor to a known compound
```

```
r4 <- getchainseq(spmeinvivo, c(162.0528, -18.0106), mass = 286.3101)

# Known-pathway screening against the built-in chain database
data(pmdchain)
r5 <- getchainseq(spmeinvivo, db = pmdchain)
r5$chainsearch

## End(Not run)
```

---

`getcorpseudospectrum` *Get Pseudo-Spectrum as peaks cluster based on correlation analysis.*

---

## Description

Get Pseudo-Spectrum as peaks cluster based on correlation analysis.

## Usage

```
getcorpseudospectrum(list, corcutoff = 0.9, rtcutoff = 10, accuracy = 4)
```

## Arguments

| | |
|---|---|
| `list` | a list with peaks intensity |
| `corcutoff` | cutoff of the correlation coefficient, default 0.9 |
| `rtcutoff` | cutoff of the distances in cluster, default 10 |
| `accuracy` | measured mass or mass to charge ratio in digits, default 4 |

## Value

list with Pseudo-Spectrum index

## Examples

```
data(spmeinvivo)
pseudo <- getcorpseudospectrum(spmeinvivo)
```

gethomolog *Find homologous series by a repeating PMD unit*

## Description

Convenience wrapper around getchainseq for the common case of detecting homologous series: chains of features linked by repeated applications of a single PMD unit. Useful for finding alkyl chain series (+CH2 = 14.0157), PEG series (+C2H4O = 44.0262), polymeric artifacts, and similar patterns.

## Usage

```
gethomolog(list, unit = 14.0157, min_len = 3L, max_len = Inf, ...)
```

## Arguments

| | |
|---|---|
| list | a pmd-style list with mz, rt (optional), data |
| unit | the repeating PMD, default 14.0157 (CH2) |
| min_len | minimum number of nodes in the series, default 3 |
| max_len | maximum number of nodes, default Inf |
| ... | passed through to getchainseq (e.g. corcutoff, rtdir, mass, ppm) |

## Value

as getchainseq

## See Also

getchainseq

## Examples

```
## Not run:
data(spmeinvivo)
gethomolog(spmeinvivo, unit = 14.0157, min_len = 4)
gethomolog(spmeinvivo, unit = 44.0262, min_len = 3)

## End(Not run)
```

## getms2pmd *read in MSP file as list for ms/ms annotation*

### Description

read in MSP file as list for ms/ms annotation

### Usage

```
getms2pmd(file, digits = 2, icf = 10)
```

### Arguments

| | |
|---|---|
| file | the path to your MSP file |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| icf | intensity cutoff, default 10 percentage |

### Value

list a list with MSP information for MS/MS annotation

## getmspmd *read in MSP file as list for EI-MS annotation*

### Description

read in MSP file as list for EI-MS annotation

### Usage

```
getmspmd(file, digits = 2, icf = 10)
```

### Arguments

| | |
|---|---|
| file | the path to your MSP file |
| digits | mass or mass to charge ratio accuracy for pmd, default 0 |
| icf | intensity cutoff, default 10 percentage |

### Value

list a list with MSP information for EI-MS annotation

getnse *Calculate Normalized Spectral Entropy (NSE)*

## Description

Calculate Normalized Spectral Entropy (NSE)

## Usage

```
getnse(x)
```

## Arguments

| | |
|---|---|
| x | Matrix or dataframe with samples in column and features in row |

## Value

Normalized Spectral Entropy (NSE) between 0 and 1

## Examples

```
data(spmeinvivo)
getnse(spmeinvivo$data)
```

getpaired *Filter ions/peaks based on retention time hierarchical clustering, paired mass distances(PMD) and PMD frequency analysis.*

## Description

Filter ions/peaks based on retention time hierarchical clustering, paired mass distances(PMD) and PMD frequency analysis.

## Usage

```
getpaired(
  list,
  rtcutoff = 10,
  ng = NULL,
  digits = 2,
  accuracy = 4,
  corcutoff = NULL
)
```

### Arguments

| | |
|---|---|
| list | a peaks list with mass to charge, retention time and intensity data |
| rtcutoff | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| ng | cutoff of global PMD's retention time group numbers, If ng = NULL, 20 percent of RT cluster will be used as ng, default NULL. |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |
| corcutoff | cutoff of the correlation coefficient, 0.6 is suggested, default NULL |

### Value

list with tentative isotope, multi-chargers, adducts, and neutral loss peaks' index, retention time clusters.

### See Also

getstd,getsda,plotpaired

### Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
```

---

## getpmd *Get pmd for specific reaction*

---

### Description

Get pmd for specific reaction

### Usage

```
getpmd(list, pmd, rtcutoff = 10, corcutoff = NULL, digits = 2, accuracy = 4)
```

### Arguments

| | |
|---|---|
| list | a list with mzrt profile |
| pmd | a specific paired mass distance or a vector of pmds |
| rtcutoff | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| corcutoff | cutoff of the correlation coefficient, default NULL |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |

## Value

list with paired peaks for specific pmd or pmds.

## See Also

getpaired,getstd,getsda,getrda

## Examples

```
data(spmeinvivo)
pmd <- getpmd(spmeinvivo,pmd=15.99)
```

---

getpmddf *Get pmd details for specific reaction after the removal of isotopouge.*

---

## Description

Get pmd details for specific reaction after the removal of isotopouge.

## Usage

```
getpmddf(mz, group = NULL, pmd = NULL, digits = 2, mdrange = c(0.25, 0.9))
```

## Arguments

| | |
|---|---|
| mz | a vector of mass to charge ratio. |
| group | mass to charge ratio group from either retention time or mass spectrometry imaging segmentation. |
| pmd | a specific paired mass distance or a vector of pmds |
| digits | mass or mass to charge ratio accuracy for pmd, default 2. |
| mdrange | mass defect range to ignore. Default c(0.25,0.9) to retain the possible reaction related paired mass. |

## Value

dataframe with paired peaks for specific pmd or pmds. When group is provided, a column named net will be generated to show if certain pmd will be local(within the same group) or global(across the groups)

## See Also

getpaired,getstd,getsda,getrda

## Examples

```
data(spmeinvivo)
pmddf <- getpmddf(spmeinvivo$mz,pmd=15.99)
```

`getposneg` *Link pos mode peak list with neg mode peak list by pmd.*

## Description

Link pos mode peak list with neg mode peak list by pmd.

## Usage

```
getposneg(pos, neg, pmd = 2.02, digits = 2)
```

## Arguments

| | |
|---|---|
| `pos` | a list with mzrt profile collected from positive mode. |
| `neg` | a list with mzrt profile collected from negative mode. |
| `pmd` | numeric or numeric vector |
| `digits` | mass or mass to charge ratio accuracy for pmd, default 2 |

## Value

dataframe with filtered positive and negative peak list

`getpseudospectrum` *Get Pseudo-Spectrum as peaks cluster based on correlation analysis.*

## Description

Get Pseudo-Spectrum as peaks cluster based on correlation analysis.

## Usage

```
getpseudospectrum(
  list,
  corcutoff = NULL,
  rtcutoff = 10,
  accuracy = 4,
  ng = NULL,
  digits = 2
)
```

### Arguments

| | |
|---|---|
| list | A list containing peak intensities, m/z values, and retention times. Must include elements: mz, rt, data. |
| corcutoff | Cutoff value for correlation coefficient (default: 0.9). |
| rtcutoff | Cutoff value for retention time clustering (default: 10). |
| accuracy | Number of decimal places for m/z rounding (default: 4). |
| ng | cutoff of global PMD's retention time group numbers, If ng = NULL, 20 percent of RT cluster will be used as ng, default NULL. |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |

### Value

A list with pseudo-spectrum clustering results.

### Examples

```
data(spmeinvivo)
pseudo <- getpseudospectrum(spmeinvivo)
```

---

## getrda *Perform structure/reaction directed analysis for mass only.*

---

### Description

Perform structure/reaction directed analysis for mass only.

### Usage

```
getrda(
  mz,
  pmd = NULL,
  freqcutoff = 10,
  digits = 3,
  top = 20,
  formula = NULL,
  mdrange = c(0.25, 0.9),
  verbose = FALSE
)
```

### Arguments

| | |
|---|---|
| mz | numeric vector for independent mass or mass to charge ratio. Mass to charge ratio from GlobalStd algorithm is suggested. Isomers would be excluded automated |
| pmd | a specific paired mass distance or a vector of pmds, default NULL |

| | |
|---|---|
| freqcutoff | pmd frequency cutoff for structures or reactions, default 10 |
| digits | mass or mass to charge ratio accuracy for pmd, default 3 |
| top | top n pmd frequency cutoff when the freqcutoff is too small for large data set |
| formula | vector for formula when you don't have mass or mass to charge ratio data |
| mdrange | mass defect range to ignore. Default c(0.25,0.9) to retain the possible reaction related paired mass |
| verbose | logic, if TURE, return will be llist with paired mass distances table. Default FALSE. |

## Value

logical matrix with row as the same order of mz or formula and column as high frequency pmd group when verbose is FALSE

## See Also

getsda

## Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
sda <- getrda(spmeinvivo$mz[std$stdmassindex])
sda <- getrda(spmeinvivo$mz, pmd = c(2.016,15.995,18.011,14.016))
```

---

# getreact *Get quantitative paired peaks list for specific reaction/pmd*

---

## Description

Get quantitative paired peaks list for specific reaction/pmd

## Usage

```
getreact(
  list,
  pmd,
  rtcutoff = 10,
  digits = 2,
  accuracy = 4,
  cvcutoff = 30,
  outlier = FALSE,
  method = "static",
  ...
)
```

### Arguments

| | |
|---|---|
| list | a list with mzrt profile and data |
| pmd | a specific paired mass distances |
| rtcutoff | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |
| cvcutoff | ratio or intensity cv cutoff for quantitative paired peaks, default 30 |
| outlier | logical, if true, outlier of ratio will be removed, default False. |
| method | quantification method can be 'static' or 'dynamic'. See details. |
| ... | other parameters for getpmd |

### Details

PMD based reaction quantification methods have two options: 'static' will only consider the stable mass pairs across samples and such reactions will be limited by the enzyme or other factors than substrates. 'dynamic' will consider the unstable paired masses by normalization the relatively unstable peak with stable peak between paired masses and such reactions will be limited by one or both peaks in the paired masses.

### Value

list with quantitative paired peaks.

### See Also

getpaired,getstd,getsda,getrda,getpmd,

### Examples

```
data(spmeinvivo)
pmd <- getreact(spmeinvivo,pmd=15.99)
```

---

## getsda *Perform structure/reaction directed analysis for peaks list.*

---

### Description

Perform structure/reaction directed analysis for peaks list.

## Usage

```
getsda(
  list,
  rtcutoff = 10,
  corcutoff = NULL,
  digits = 2,
  accuracy = 4,
  freqcutoff = NULL
)
```

## Arguments

| | |
|---|---|
| list | a list with mzrt profile |
| rtcutoff | cutoff of the distances in retention time hierarchical clustering analysis, default 10 |
| corcutoff | cutoff of the correlation coefficient, default NULL |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |
| freqcutoff | pmd frequency cutoff for structures or reactions, default NULL. This cutoff will be found by PMD network analysis when it is NULL. |

## Value

list with tentative isotope, adducts, and neutral loss peaks' index, retention time clusters.

## See Also

getpaired,getstd,plotpaired

## Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
sda <- getsda(std)
```

---

# getsim *Evaluate the similarity of two matrices or dataframes*

---

## Description

Evaluate the similarity of two matrices or dataframes

## Usage

```
getsim(x, y, dim = NULL)
```

### Arguments

| | |
|---|---|
| x | Matrix or dataframe with samples in column and features in row (Original data) |
| y | Matrix or dataframe with samples in column and features in row (Reduced data) |
| dim | number of retained dimensions for PCASF. Defaults to all. |

### Value

A numeric vector containing PCASF, RV coefficient, and Mantel test r (Pearson and Spearman).

### Examples

```
data(spmeinvivo)
re <- globalstd(spmeinvivo)
x <- spmeinvivo$data
y <- spmeinvivo$data[re$stdmassindex, ]
getsim(x, y)
```

---

## getstd *Identify standard ions through retention time clustering and PMD relationships*

---

### Description

Identify standard ions through retention time clustering and PMD relationships

### Usage

```
getstd(list, digits = 2, accuracy = 4)
```

### Arguments

| | |
|---|---|
| list | A list object from getpaired() containing paired features |
| digits | Rounding digits for mass differences |
| accuracy | Mass accuracy for standard ion identification |

### Value

List with added standard ion indices and metadata

### See Also

getpaired,getsda,plotstd

### Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
```

## gettarget *Get multiple injections index for selected retention time*

### Description

Get multiple injections index for selected retention time

### Usage

```
gettarget(rt, drt = 10, n = 6)
```

### Arguments

| | |
|---|---|
| rt | retention time vector for peaks in seconds |
| drt | retention time drift for targeted analysis in seconds, default 10. |
| n | max ions numbers within retention time drift windows |

### Value

index for each injection

### Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
index <- gettarget(std$rt[std$stdmassindex])
table(index)
```

## globalstd *GlobalStd algorithm with structure/reaction directed analysis*

### Description

GlobalStd algorithm with structure/reaction directed analysis

### Usage

```
globalstd(
  list,
  rtcutoff = 10,
  ng = NULL,
  corcutoff = NULL,
  digits = 2,
  accuracy = 4,
  freqcutoff = NULL,
  sda = FALSE
)
```

### Arguments

| | |
|---|---|
| list | a peaks list with mass to charge, retention time and intensity data |
| rtcutoff | cutoff of the distances in cluster, default 10 |
| ng | cutoff of global PMD's retention time group numbers, If ng = NULL, 20 percent of RT cluster will be used as ng, default NULL. |
| corcutoff | cutoff of the correlation coefficient, default NULL |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| accuracy | measured mass or mass to charge ratio in digits, default 4 |
| freqcutoff | pmd frequency cutoff for structures or reactions, default NULL. This cutoff will be found by PMD network analysis when it is NULL. |
| sda | logical, option to perform structure/reaction directed analysis, default FALSE. |

### Value

list with GlobalStd algorithm processed data.

### See Also

getpaired,getstd,getsda,plotstd,plotstdsda,plotstdrt

### Examples

```
data(spmeinvivo)
re <- globalstd(spmeinvivo)
```

---

## hmdb

*A dataframe containing HMDB with unique accurate mass pmd with three digits frequency larger than 1 and accuracy percentage larger than 0.9.*

---

### Description

A dataframe containing HMDB with unique accurate mass pmd with three digits frequency larger than 1 and accuracy percentage larger than 0.9.

### Usage

```
data(hmdb)
```

### Format

A dataframe with atoms numbers of C, H, O, N, P, S

**percentage** accuracy of atom numbers prediction

**pmd2** pmd with two digits

**pmd** pmd with three digits

## keggrall — *A dataframe containing reaction related accurate mass pmd and related reaction formula with KEGG ID*

### Description

A dataframe containing reaction related accurate mass pmd and related reaction formula with KEGG ID

### Usage

```
data(keggrall)
```

### Format

A dataframe with KEGG reaction, their realted pmd and atoms numbers of C, H, O, N, P, S

**ID** KEGG reaction ID

**pmd** pmd with three digits

## MaConDa — *mass spectrometry contaminants database for PMD check*

### Description

mass spectrometry contaminants database for PMD check

### Usage

```
data(MaConDa)
```

### Format

A data frame from doi:10.1093/bioinformatics/bts527 with 308 rows and 5 variables:

**id** MaConDa ID

**name** contaminants

**formula** contaminants fomula

**exact_mass** exact mass of contaminants

**type_of_contaminant** type of contaminant

`omics` *A dataframe containing multiple reaction database ID and their related accurate mass pmd and related reactions*

## Description

A dataframe containing multiple reaction database ID and their related accurate mass pmd and related reactions

## Usage

```
data(omics)
```

## Format

A dataframe with reaction and their realted pmd

**KEGG** KEGG reaction ID

**RHEA_ID** RHEA_ID

**DIRECTION** reaction direction

**MASTER_ID** master reaction RHEA ID

**ec** ec reaction ID

**ecocyc** ecocyc reaction ID

**macie** macie reaction ID

**metacyc** metacyc reaction ID

**reactome** reactome reaction ID

**compounds** reaction related compounds

**pmd** pmd with two digits

**pmd2** pmd with three digits

`parse_pmd_pattern` *Parse a PMD pattern string into a list of step specs*

## Description

Converts a compact string grammar into the list-of-steps format accepted by `getchainseq`. Steps are separated by semicolons. Each step is a numeric PMD (or `*` for wildcard) optionally followed by a quantifier:

- `+` one or more (min=1, max=Inf)
- `*` zero or more (min=0, max=Inf)
- `?` zero or one (min=0, max=1)

- {n} exactly n
- {n,m} n to m times
- {n,} at least n times

Whitespace is ignored.

## Usage

```
parse_pmd_pattern(s)
```

## Arguments

| | |
|---|---|
| s | a single character string, e.g. "162.0528; -18.0106{0,3}; 14.0157+" |

## Value

a list of step specs suitable for getchainseq(pattern = ...)

## Examples

```
parse_pmd_pattern("162.0528; -18.0106{0,3}")
parse_pmd_pattern("*; 14.0157+")
```

---

# pcasf *Compare matrices using PCA similarity factor*

---

## Description

Compare matrices using PCA similarity factor

## Usage

```
pcasf(x, y, dim = NULL)
```

## Arguments

| | |
|---|---|
| x | Matrix with sample in column and features in row |
| y | Matrix is compared to x. |
| dim | number of retained dimensions in the comparison. Defaults to all. |

## Value

Ratio of projected variance to total variance

## Author(s)

Edgar Zanella Alvarenga

## References

Singhal, A. and Seborg, D. E. (2005), Clustering multivariate time-series data. J. Chemometrics, 19: 427-438. doi: 10.1002/cem.945

## Examples

```
c1 <- matrix(rnorm(16),nrow=4)
c2 <- matrix(rnorm(16),nrow=4)
pcasf(c1, c2)
```

---

`plotcn` *plot PMD KEGG network for certain compounds and output network average distance and degree*

---

## Description

plot PMD KEGG network for certain compounds and output network average distance and degree

## Usage

```
plotcn(formula, name, pmd)
```

## Arguments

| | |
|---|---|
| `formula` | Chemical formula |
| `name` | Compound name |
| `pmd` | specific paired mass distances |

## Examples

```
plotcn('C6H12O6','Glucose',c(2.016,14.016,15.995))
```

---

`plotpaired` *Plot the mass pairs and high frequency mass distances*

---

## Description

Plot the mass pairs and high frequency mass distances

## Usage

```
plotpaired(list, index = NULL, digits = 2, ...)
```

### Arguments

| | |
|---|---|
| list | a list from getpaired function |
| index | index for PMD value |
| digits | mass or mass to charge ratio accuracy for pmd, default 2 |
| ... | other parameters for plot function |

### See Also

getpaired, globalstd

### Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
plotpaired(pmd)
```

---

## plotrtg *Plot the retention time group*

---

### Description

Plot the retention time group

### Usage

```
plotrtg(list, ...)
```

### Arguments

| | |
|---|---|
| list | a list from getpaired function |
| ... | other parameters for plot function |

### See Also

getpaired, globalstd

### Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
plotrtg(pmd)
```

## plotsda *Plot the specific structure directed analysis(SDA) groups*

### Description

Plot the specific structure directed analysis(SDA) groups

### Usage

```
plotsda(list, ...)
```

### Arguments

| | |
|---|---|
| list | a list from getpmd function |
| ... | other parameters for plot function |

### See Also

getstd, globalstd,plotstd,plotpaired,plotstdrt

### Examples

```
data(spmeinvivo)
re <- getpmd(spmeinvivo,pmd=78.9)
plotsda(re)
```

## plotstd *Plot the std mass from GlobalStd algorithm*

### Description

Plot the std mass from GlobalStd algorithm

### Usage

```
plotstd(list)
```

### Arguments

| | |
|---|---|
| list | a list from getstd function |

### See Also

getstd, globalstd

## Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
plotstd(std)
```

---

`plotstdrt` *Plot the std mass from GlobalStd algorithm in certain retention time groups*

---

## Description

Plot the std mass from GlobalStd algorithm in certain retention time groups

## Usage

```
plotstdrt(list, rtcluster, ...)
```

## Arguments

| | |
|---|---|
| `list` | a list from getstd function |
| `rtcluster` | retention time group index |
| `...` | other parameters for plot function |

## See Also

`getstd`, `globalstd`,`plotstd`,`plotpaired`,`plotstdsda`

## Examples

```
data(spmeinvivo)
pmd <- getpaired(spmeinvivo)
std <- getstd(pmd)
plotstdrt(std,rtcluster = 6)
```

## plotstdsda *Plot the std mass from GlobalStd algorithm in structure directed analysis(SDA) groups*

### Description

Plot the std mass from GlobalStd algorithm in structure directed analysis(SDA) groups

### Usage

```
plotstdsda(list, index = NULL, ...)
```

### Arguments

| | |
|---|---|
| list | a list from getsda function |
| index | index for PMD value |
| ... | other parameters for plot function |

### See Also

getstd, globalstd,plotstd,plotpaired,plotstdrt

### Examples

```
data(spmeinvivo)
re <- globalstd(spmeinvivo, sda=TRUE)
plotstdsda(re)
```

## pmdchain *Curated database of known multi-step reaction chains*

### Description

A hand-curated set of named biotransformation chains, each an ordered sequence of paired mass distances (PMDs) for known-pathway screening of untargeted feature lists via `getchainseq(list, db = pmdchain)`. Each chain spans several rows (one per step). Every step PMD is the exact monoisotopic mass of its CHNOPS element delta. `keggcount` is an observable specificity score: the number of compound-connected KEGG reaction paths (from `keggrall`) whose PMD signature matches the whole chain – lower means a rarer, more specific signature (xenobiotic phase-II conjugations are most specific; common edits such as methylation/phosphorylation are highly degenerate). A chain match is a relational annotation, not compound identification.

### Usage

```
data(pmdchain)
```

### Format

A data frame with one row per chain step:

**chain_id** short identifier of the chain

**name** human-readable chain name

**class** natural metabolism or xenobiotic (phase-I/II) chemistry

**nstep** number of steps in the chain

**step** 1-based step index

**transformation** the single-step transformation block name

**pmd** exact monoisotopic mass of the step's element delta (Da)

**dC, dH, dN, dO, dP, dS** signed CHNOPS element deltas of the step

**keggcount** number of KEGG reaction paths sharing the chain's PMD signature (specificity proxy; lower = rarer)

**description** biochemical description of the chain

---

## runPMD *Shiny application for PMD analysis*

---

### Description

Shiny application for PMD analysis

### Usage

```
runPMD()
```

---

## runPMDnet *Shiny application for PMD network analysis*

---

### Description

Shiny application for PMD network analysis

### Usage

```
runPMDnet()
```

## sda

*A dataset containing common Paired mass distances of substructure, ions replacements, and reaction*

### Description

A dataset containing common Paired mass distances of substructure, ions replacements, and reaction

### Usage

```
data(sda)
```

### Format

A data frame with 146 rows and 4 variables:

**PMD** Paired mass distances

**origin** potential sources

**Ref.** references

**mode** b for biological reaction and e for environmental reaction

## spmeinvivo

*A peaks list dataset containing 9 samples from 3 fish with triplicates samples for each fish from LC-MS.*

### Description

A peaks list dataset containing 9 samples from 3 fish with triplicates samples for each fish from LC-MS.

### Usage

```
data(spmeinvivo)
```

### Format

A list with 4 variables from 1459 LC-MS peaks:

**mz** mass to charge ratios

**rt** retention time

**data** intensity matrix

**group** group information

# Index


∗ **datasets**
  hmdb, 20
  keggrall, 21
  MaConDa, 21
  omics, 22
  pmdchain, 28
  sda, 30
  spmeinvivo, 30

getcda, 3
getchain, 4, *5, 6*
getchainseq, 5, *8*, *22*
getcorpseudospectrum, 7
gethomolog, *6*, 8
getms2pmd, 9
getmspmd, 9
getnse, 10
getpaired, 10, *12*, *16–18*, *20*, *25*
getpmd, 11, *16*
getpmddf, 12
getposneg, 13
getpseudospectrum, 13
getrda, *3*, *12*, 14, *16*
getreact, 15
getsda, *3*, *11, 12*, *15*, 16, *16*, *18*, *20*
getsim, 17
getstd, *11, 12*, *16, 17*, 18, *20*, *26–28*
gettarget, 19
globalstd, 19, *25–28*

hmdb, 20

keggrall, 21, *28*

MaConDa, 21

omics, 22

parse_pmd_pattern, *5, 6*, 22
pcasf, 23
plotcn, 24
plotpaired, *11*, *17*, 24, *26–28*
plotrtg, 25
plotsda, 26
plotstd, *18*, *20*, 26, *26–28*
plotstdrt, *20*, *26*, 27, *28*
plotstdsda, *20*, *27*, 28
pmdchain, 28

runPMD, 29
runPMDnet, 29

sda, 30
spmeinvivo, 30