# Package 'flagme'

October 12, 2016

· · · · · · · · · · · · · · · · · · ·
Version 1.28.0
<b>Date</b> 2015/04/06
Title Analysis of Metabolomics GC/MS Data
Author Mark Robinson <mark.robinson@imls.uzh.ch>, Riccardo Romoli <riccardo.romoli@unifi.it></riccardo.romoli@unifi.it></mark.robinson@imls.uzh.ch>
Maintainer Mark Robinson <mark.robinson@imls.uzh.ch>, Riccardo Romoli <riccardo.romoli@unifi.it></riccardo.romoli@unifi.it></mark.robinson@imls.uzh.ch>
Depends gcspikelite, xcms, CAMERA
Imports gplots, graphics, MASS, methods, SparseM, stats, utils
<b>Description</b> Fragment-level analysis of gas chromatography - mass spectrometry metabolomics data
License LGPL (>= 2)
Collate Oclasses.R clusterAlignment.R init.R multipleAlignment.R peaksAlignment.R progressiveAlignment.R betweenAlignment.R dp.R metrics.R parse.R peaksDataset.R gatherInfo.R plotFragments.R rmaFitUnit.R addXCMSPeaks.R correlationAlignment.R retFatMatrix.R plotSpectra.R exportSpectra.R
biocViews DifferentialExpression, MassSpectrometry
NeedsCompilation yes
R topics documented:
addAMDISPeaks addChromaTOFPeaks addXCMSPeaks betweenAlignment calcTimeDiffs clusterAlignment compress correlationAlignment 1 correlationAlignment-class dp
dp

2 addAMDISPeaks

Index		35
	rmaFitUnit	33
	retFatMatrix	32
	progressiveAlignment-class	31
	plotSpectra	29
	plotImage	28
	plot.peaksDataset	26
	peaksDataset	25
	peaksAlignment-class	24
	parseELU	22
	parseChromaTOF	21
	normDotProduct	20
	multipleAlignment-class	18
	imputePeaks	17
	gatherInfo	15
	exportSpectra	15
	eitherMatrix-class	14

addAMDISPeaks

Add AMDIS peak detection results

## **Description**

Reads ASCII ELU-format files (output from AMDIS) and attaches them to an already created peaksDataset object

## Usage

```
addAMDISPeaks(object,fns=dir(,"[Eu][L1][Uu]"),verbose=TRUE,...)
```

## **Arguments**

object a peaksDataset object.

fns character vector of same length as object@rawdata (user ensures the order

matches)

verbose whether to give verbose output, default TRUE

... arguments passed on to parseELU

# **Details**

Repeated calls to parseELU to add peak detection results to the original peaksDataset object.

# Value

peaksDataset object

addChromaTOFPeaks 3

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
parseELU, peaksDataset
```

#### **Examples**

```
# need access to CDF (raw data) and ELU files
require(gcspikelite)
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")

# full paths to file names
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# create a 'peaksDataset' object and add AMDIS peaks to it
pd<-peaksDataset(cdfFiles[1],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1])</pre>
```

addChromaTOFPeaks

Add ChromaTOF peak detection results

#### Description

Reads ASCII tab-delimited format files (output from ChromaTOF) and attaches them to an already created peaksDataset object

#### Usage

```
addChromaTOFPeaks(object,fns=dir(,"[Tt][Xx][Tx]"),rtDivide=60,verbose=TRUE,...)
```

# Arguments

object a peaksDataset object.

fns character vector of same length as object@rawdata (user ensures the order matches)

rtDivide number giving the amount to divide the retention times by.

verbose whether to give verbose output, default TRUE

arguments passed on to parseChromaTOF

4 addXCMSPeaks

#### **Details**

Repeated calls to parseChromaTOF to add peak detection results to the original peaksDataset object.

## Value

```
peaksDataset object
```

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
parseChromaTOF, peaksDataset
```

#### **Examples**

```
# need access to CDF (raw data) and ChromaTOF files
require(gcspikelite)
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")

# full paths to file names
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
# [not run] cTofFiles<-dir(gcmsPath,"txt",full=TRUE)

# create a 'peaksDataset' object and add ChromaTOF peaks to it
pd<-peaksDataset(cdfFiles[1],mz=seq(50,550),rtrange=c(7.5,8.5))
# [not run] pd<-addChromTOFPeaks(pd,...)</pre>
```

addXCMSPeaks

Add xcms/CAMERA peak detection results

# Description

Reads the raw data using xcms, group each extracted ion according to their retention time using CAMERA and attaches them to an already created peaksDataset object

## Usage

```
addXCMSPeaks(files, object, peakPicking=c('cwt', 'mF'), ...)
```

addXCMSPeaks 5

## **Arguments**

files character vector of same length as object@rawdata (user ensures the order matches)

object a peaksDataset object.

peakPicking Methods to use for peak detection. See details.
... arguments passed on to xcmsSet and annotate

#### **Details**

Repeated calls to xcmsSet and annotate to perform peak-picking and deconvolution. The peak detection results are added to the original peaksDataset object. Two peak detection alorithms are available: continuous wavelet transform (peakPicking=c('cwt')) and the matched filter approach (peakPicking=c('mF')) described by Smith et al (2006). For further information consult the xcms package manual.

#### Value

peaksDataset object

#### Author(s)

Riccardo Romoli < riccardo . romoli@unifi.it >

#### See Also

peaksDataset findPeaks.matchedFilter findPeaks.centWave xcmsRaw-class

# **Examples**

6 betweenAlignment

betweenAlignment	Data Structure for "between" alignment of many GCMS samples

# **Description**

This function creates a "between" alignment (i.e. comparing merged peaks)

# Usage

 $between \verb|Alignment(pD,cAList,pAList,impList,filterMin=3,gap=0.7,D=10,usePeaks=TRUE,df=30,verbose=TRUE,df=$ 

## **Arguments**

pD	a peaksDataset object
cAList	list of clusterAlignment objects, one for each experimental group
pAList	list of progressiveAlignment objects, one for each experimental group
impList	list of imputation lists
filterMin	minimum number of peaks within a merged peak to be kept in the analysis
gap	gap parameter
D	retention time penalty parameter
usePeaks	logical, whether to use peaks (if TRUE) or the full 2D profile alignment (if FALSE)
df	distance from diagonal to calculate similarity
verbose	logical, whether to print information

## **Details**

betweenAlignment objects gives the data structure which stores the result of an alignment across several "pseudo" datasets. These pseudo datasets are constructed by merging the "within" alignments.

# Value

betweenAlignment object

## Author(s)

Mark Robinson

# References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

multipleAlignment

calcTimeDiffs 7

#### **Examples**

```
require(gcspikelite)
# see 'multipleAlignment'
```

calcTimeDiffs

Calculate retention time shifts from profile alignments

# Description

This function takes the set of all pairwise profile alignments and use these to estimate retention time shifts between each pair of samples. These will then be used to normalize the retention time penalty of the signal peak alignment.

## Usage

```
calcTimeDiffs(pd,ca.full,verbose=TRUE)
```

## **Arguments**

pd a peaksDataset object

ca.full a clusterAlignment object, fit with verbose logical, whether to print out information

# **Details**

Using the set of profile alignments,

## Value

list of same length as ca.full@alignments with the matrices giving the retention time penalties.

# Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

# See Also

```
peaksAlignment, clusterAlignment
```

8 clusterAlignment

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath <- paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles <- dir(gcmsPath,"CDF",full=TRUE)
eluFiles <- dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd <- peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd <- addAMDISPeaks(pd,eluFiles[1:2])

# pairwise alignment using all scans
fullca <- clusterAlignment(pd, usePeaks=FALSE, df=100)

# calculate retention time shifts
timedf <- calcTimeDiffs(pd, fullca)</pre>
```

clusterAlignment Data Structure for a collection of all pairwise alignments of GCMS runs

## Description

Store the raw data and optionally, information regarding signal peaks for a number of GCMS runs

#### Usage

```
cluster A lignment (pD, runs=1: length (pD@rawdata), timed f=NULL, use Peaks=TRUE, verbose=TRUE, \dots)
```

# Arguments

pD	a peaksDataset object.
runs	vector of integers giving the samples to calculate set of pairwise alignments over.
timedf	list (length = the number of pairwise alignments) of matrices giving the expected time differences expected at each pair of peaks (used with usePeaks=TRUE, passed to peaksAlignment
usePeaks	logical, TRUE uses peakdata list, FALSE uses rawdata list for computing similarity.
verbose	logical, whether to print out info.
	other arguments passed to peaksAlignment

# **Details**

clusterAlignment computes the set of pairwise alignments.

compress 9

#### Value

clusterAlignment object

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
peaksDataset, peaksAlignment
```

# **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])
ca<-clusterAlignment(pd, gap = .5,D=.05,df=30)</pre>
```

compress

Compress an alignment object

# Description

Many of the peaks are not similar. So, the set of pairwise similarity matrices can be compressed.

## Usage

```
compress(object,verbose=TRUE,...)
decompress(object,verbose=TRUE,...)
```

#### Arguments

object a peaksAlignment, peaksAlignment or peaksAlignment object to be compressed

verbose logical, whether to print out information

... further arguments

correlation Alignment

#### **Details**

Using sparse matrix representations, a significant compression can be achieved. Here, we use the matrix.csc class of the SpareM package.

#### Value

an object of the same type as the input object

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
peaksAlignment, clusterAlignment, progressiveAlignment
```

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])

# pairwise alignment (it is compressed by default)
ca<-clusterAlignment(pd, usePeaks = TRUE, df = 20)
object.size(ca)

# decompress
ca<-decompress(ca)
object.size(ca)</pre>
```

 $\begin{tabular}{ll} correlation A lignment & Function for correlation-based a lignment strategy of multiple GCMS \\ & samples \end{tabular}$ 

correlationAlignment 11

#### **Description**

Performs a correlation-based peak alignment of multiple GCMS peak lists; this function is able to align multiple samples, by a center-star strategy.

# Usage

```
correlationAlignment(object, thr=0.85, D=20, penality=0.2, normalize=TRUE, minFilter=1)
```

#### **Arguments**

object a peaksDataset object

thr correlation trashold from 0 (min) to 1 (max)

D retention time window in seconds

penality the penality inflicted to a match between two peaks when the retention time

difference exceed the parameter D

normalize logical, whether to use normalized-to-100 peaks intensity or as such

minFilter if a feature is matched in a number of samples less than minFilter, this feature

is trashed. The value of minFilter must be smaller than the number of samples

#### **Details**

The correlation-based peak alignment for multiple GCMS peak lists uses a center-star tecnique to the alignment of the peaks. The combination of the D and penality parameters allow the users to force the algorithm to match the peaks close to the reference. The thr parameter control the matching factor.

# Value

```
correlationAlignment object
```

#### Author(s)

Riccardo Romoli <riccardo.romoli@unifi.it>

#### See Also

```
peaksDataset, addXCMSPeaks, correlationAlignment-class
```

## **Examples**

```
require(gcspikelite)
# paths and files
gcmsPath <- paste(find.package("gcspikelite"), "data", sep="/")
cdfFiles <- dir(gcmsPath, "CDF", full=TRUE)

# read data, peak detection results
pd <- peaksDataset(cdfFiles[1:2], mz=seq(50,550), rtrange=c(7.5,8.5))
pd <- addXCMSPeaks(files=cdfFiles[1:2], object=pd, peakPicking=c('mF'),</pre>
```

correlationAlignment-class

"correlationAlignment-class"

#### **Description**

A class containing the index of the aligned chromatographic peaks.

## **Objects from the Class**

Objects can be created by calls of the form new("correlationAlignment", ...). The object created contains both the results of the alignment procedure and the file used and as center-star.

#### **Slots**

Alignment: Object of class "align" contain the matrix of the aligned features. The rows represent the different peaks while the columns represent the files. The values of the matrix refers to peaksind slot of the peaksDataSet object.

Center: Object of class "character" contain the file used as a center-star.

#### Methods

```
show signature(object = "correlationAlignment")
```

# Author(s)

Riccardo Romoli <riccardo.romoli@unifi.it>

#### See Also

 ${\tt correlationAlignment}\ for\ further\ information\ about\ the\ alignment\ function.$ 

#### **Examples**

```
showClass("correlationAlignment")
```

*dp* 13

dp Dynamic programming algorithm, given a similarity matrix

# Description

This function calls C code for a bare-bones dynamic programming algorithm, finding the best cost path through a similarity matrix.

## Usage

```
dp(M,gap=.5,big=1000000000,verbose=FALSE)
```

#### **Arguments**

M similarity matrix
gap penalty for gaps
big large value used for matrix margins

logical, whether to print out information

# Details

verbose

This is a pretty standard implementation of a bare-bones dynamic programming algorithm, with a single gap parameter and allowing only simple jumps through the matrix (up, right or diagonal).

#### Value

list with element match with the set of pairwise matches.

#### Author(s)

Mark Robinson

## References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

normDotProduct

14 eitherMatrix-class

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])

# similarity matrix
r<-normDotProduct(pd@peaksdata[[1]],pd@peaksdata[[2]])

# dynamic-programming-based matching of peaks
v<-dp(r,gap=.5)</pre>
```

eitherMatrix-class

The eitherMatrix class

## **Description**

A container to store either matrix or matrix.csc objects

# Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

## See Also

peaksAlignment

exportSpectra 15

exportSpectra	
exportSpectra	

#### **Description**

Write the deconvoluted mass spectra to an external file

## Usage

```
exportSpectra(object, sample, spectraID, normalize = TRUE)
```

# **Arguments**

object an object of class "peaksDataset" where to keep the mass spectra; both abun-

dance (y) than m/z(x)

sample character, the sample from were to plot the mass spectra

spectraID numerical, a vector containing the index of the spectra to be plotted.

normalize logical, if TRUE normalize the intensity of the mass peak to 100, the most abun-

dant is 100 other peaks are scaled consequetially

#### **Details**

Write a .msp file of the deconvoluted mass spectra. Usfull to try to identify the unknown spectra using NIST Search.

# Value

a .msp file ready to be read using NIST search

#### Author(s)

riccardo.romoli@unifi.it

gatherInfo	Gathers abundance informations from an alignment

## **Description**

Given an alignment table (indices of matched peaks across several samples) such as that within a progressiveAlignment or multipleAlignment object, this routines goes through the raw data and collects the abundance of each fragment peak, as well as the retention times across the samples.

#### Usage

```
gatherInfo(pD, obj, newind = NULL, method = c("apex"), findmzind = TRUE, useTIC = FALSE, top = NULL, in
```

16 gatherInfo

#### **Arguments**

pD a peaksDataset object, to get the abundance data from

obj either a multipleAlignment or progressiveAlignment object

newind list giving the

method method used to gather abundance information, only apex implemented cur-

rently.

findmzind logical, whether to take a subset of all m/z indices

useTIC logical, whether to use total ion current for abundance summaries

top only use the top top peaks

intensity.cut percentage of the maximum intensity

#### **Details**

This procedure loops through the table of matched peaks and gathers the

#### Value

Returns a list (of lists) for each row in the alignment table. Each list has 3 elements:

mz a numerical vector of the m/z fragments used

rt a numerical vector for the exact retention time of each peak across all samples data matrix of fragment intensities. If useTIC=TRUE, this matrix will have a single

row

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

imputePeaks

## **Examples**

```
require(gcspikelite)
# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)
# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))</pre>
```

imputePeaks 17

```
pd<-addAMDISPeaks(pd,eluFiles[1:2])

# multiple alignment
ma<-multipleAlignment(pd,c(1,1),wn.gap=0.5,wn.D=.05,bw.gap=0.6,bw.D=.2,usePeaks=TRUE,filterMin=1,df=50,verbose

# gather apex intensities
d<-gatherInfo(pd,ma)

# table of retention times
nm<-list(paste("MP",1:length(d),sep=""),c("S1","S2"))
rts<-matrix(unlist(sapply(d,.subset,"rt")),byrow=TRUE,nc=2,dimnames=nm)</pre>
```

imputePeaks

Imputatin of locations of peaks that were undetected

#### **Description**

Using the information within the peaks that are matched across several runs, we can impute the location of the peaks that are undetected in a subset of runs

#### Usage

```
imputePeaks(pD, obj, type = 1, obj2 = NULL, filterMin = 3, verbose = TRUE)
```

# Arguments

pD	a peaksDataset object
obj	the alignment object, either multipleAlignment or progressiveAlignment, that is used to infer the unmatched peak locations
type	type of imputation to do, 1 for simple linear interpolation (default), 2 only works if obj2 is a clusterAlignment object
obj2	a clusterAlignment object
filterMin	minimum number of peaks within a merged peak to impute
verbose	logical, whether to print out information

# Details

If you are aligning several samples and for a (small) subset of the samples in question, a peak is undetected, there is information within the alignment that can be useful in determining where the undetected peak is, based on the surrounding matched peaks. Instead of moving forward with missing values into the data matrices, this procedures goes back to the raw data and imputes the location of the apex (as well as the start and end), so that we do not need to bother with post-hoc imputation or removing data because of missing components.

We realize that imputation is prone to error and prone to attributing intensity from neighbouring peaks to the unmatched peak. We argue that this is still better than having to deal with these in statistical models after that fact. This may be an area of future improvement.

#### Value

list with 3 elements apex, start and end, each masked matrices giving the scan numbers of the imputed peaks.

## Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

multipleAlignment, progressiveAlignment, peaksDataset

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:3],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:3])

# alignments
ca<-clusterAlignment(pd, gap = .5,D=.05,df=30)
pa<-progressiveAlignment(pd, ca, gap = .6, D=.1,df=30)

v<-imputePeaks(pd,pa,filterMin=1)</pre>
```

multipleAlignment-class

Data Structure for multiple alignment of many GCMS samples

#### **Description**

Store the raw data and optionally, information regarding signal peaks for a number of GCMS runs

#### Usage

```
multipleAlignment(pd,group,bw.gap=0.8,wn.gap=0.6,bw.D=.20,wn.D=.05,filterMin=3,lite=FALSE,usePeaks=
```

#### **Arguments**

pd	a peaksDataset object
group	factor variable of experiment groups, used to guide the alignment algorithm
bw.gap	gap parameter for "between" alignments
wn.gap	gap parameter for "within" alignments
bw.D	distance penalty for "between" alignments
wn.D	distance penalty for "within" alignments
filterMin	minimum number of peaks within a merged peak to be kept in the analysis
lite	logical, whether to keep "between" alignment details (default, FALSE)
usePeaks	logical, whether to use peaks (if TRUE) or the full 2D profile alignment (if FALSE)
df	distance from diagonal to calculate similarity
verbose	logical, whether to print information
timeAdjust	logical, whether to use the full 2D profile data to estimate retention time drifts (Note: time required) $\frac{1}{2}$
doImpute	logical, whether to impute the location of unmatched peaks

#### **Details**

multipleAlignment is the data structure giving the result of an alignment across several GCMS runs.

Multiple alignments are done progressively. First, all samples with the same tg\$Group label with be aligned (denoted a "within" alignment). Second, each group will be summarized into a pseudodata set, essentially a spectrum and retention time for each matched peak of the within-alignment. Third, these "merged peaks" are aligned in the same progressive manner, here called a "between" alignment.

#### Value

multipleAlignment object

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

# See Also

peaksDataset, betweenAlignment, progressiveAlignment

20 normDotProduct

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])

# multiple alignment
ma<-multipleAlignment(pd,c(1,1),wn.gap=0.5,wn.D=.05,bw.gap=0.6,bw.D=.2,usePeaks=TRUE,filterMin=1,df=50,verbose</pre>
```

normDotProduct

Normalized Dot Product

## **Description**

This function calculates the similarity of all pairs of peaks from 2 samples, using the spectra similarity

## Usage

```
normDotProduct(x1,x2,t1=NULL,t2=NULL,df=max(ncol(x1),ncol(x2)),D=100000,timedf=NULL,verbose=FALSE)\\
```

# Arguments

x1	data matrix for sample 1
x2	data matrix for sample 2
t1	vector of retention times for sample 1
t2	vector of retention times for sample 2
df	distance from diagonal to calculate similarity
D	retention time penalty
timedf	matrix of time differences to normalize to. if NULL, $\boldsymbol{0}$ is used.
verbose	logical, whether to print out information

#### **Details**

Efficiently computes the normalized dot product between every pair of peak vectors and returns a similarity matrix. C code is called.

#### Value

matrix of similarities

parseChromaTOF 21

#### Author(s)

Mark Robinson

## References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
dp, peaksAlignment
```

# **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])
r<-normDotProduct(pd@peaksdata[[1]],pd@peaksdata[[2]])</pre>
```

parseChromaTOF

Parser for ChromaTOF files

# Description

Reads ASCII ChromaTOF-format files from AMDIS (Automated Mass Spectral Deconvolution and Identification System)

# Usage

```
parseChromaTOF(fn,min.pc=.01,mz=seq(85,500),rt.cut=.008,rtrange=NULL,skip=1,rtDivide=60)
```

## **Arguments**

fn	ChromaTOF filename to read.
min.pc	minimum percent of maximum intensity.
mz	vector of mass-to-charge bins of raw data table.
rt.cut	the difference in retention time, below which peaks are merged together.
rtrange	retention time range to parse peaks from, can speed up parsing if only interested in a small region (must be numeric vector of length 2)
skip	number of rows to skip at beginning of the ChromaTOF
rtDivide	multiplier to divide the retention times by (default: 60)

22 parseELU

#### **Details**

parseChromaTOF will typically be called by addChromaTOFPeaks, not called directly.

Peaks that are detected within rt.cut are merged together. This avoids peaks which are essentially overlapping.

Fragments that are less than min.pc of the maximum intensity fragment are discarded.

#### Value

list with components peaks (table of spectra – rows are mass-to-charge and columns are the different detected peaks) and tab (table of features for each detection), according to what is stored in the ChromaTOF file.

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

addAMDISPeaks

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
tofFiles<-dir(gcmsPath,"tof",full=TRUE)

# parse ChromaTOF file
cTofList<-parseChromaTOF(tofFiles[1])</pre>
```

parseELU

Parser for ELU files

# Description

Reads ASCII ELU-format files from AMDIS (Automated Mass Spectral Deconvolution and Identification System)

#### Usage

```
parseELU(f,min.pc=.01,mz=seq(50,550),rt.cut=.008,rtrange=NULL)
```

parseELU 23

## **Arguments**

f	ELU filename to read.

min.pc minimum percent of maximum intensity.

mz vector of mass-to-charge bins of raw data table.

rt.cut the difference in retention time, below which peaks are merged together.

rtrange retention time range to parse peaks from, can speed up parsing if only interested

in a small region (must be numeric vector of length 2)

#### Details

parseELU will typically be called by addAMDISPeaks, not called directly.

Peaks that are detected within rt.cut are merged together. This avoids peaks which are essentially overlapping.

Fragments that are less than min.pc of the maximum intensity fragment are discarded.

#### Value

list with components peaks (table of spectra – rows are mass-to-charge and columns are the different detected peaks) and tab (table of features for each detection), according to what is stored in the ELU file.

# Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

addAMDISPeaks

# **Examples**

```
require(gcspikelite)
# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)
# parse ELU file
eluList<-parseELU(eluFiles[1])</pre>
```

24 peaksAlignment-class

peaksAlignment-class Data Structure for pairwise alignment of 2 GCMS samples

## **Description**

Store the raw data and optionally, information regarding signal peaks for a number of GCMS runs

# Usage

peaks A lignment (d1, d2, t1, t2, gap=.5, D=1000, time df=NULL, df=30, verbose=TRUE, use Peaks=TRUE, compress=TRUE, df=30, verbose=TRUE, use Peaks=TRUE, compress=TRUE, df=30, verbose=TRUE, df=30,

## **Arguments**

d1	matrix of MS intensities for 1st sample (if doing a peak alignment, this contains peak apexes/areas; if doing a profile alignment, this contains scan intensities. Rows are m/z bins, columns are peaks/scans.
d2	matrix of MS intensities for 2nd sample
t1	vector of retention times for 1st sample
t2	vector of retention times for 2nd sample
gap	gap penalty for dynamic programming algorithm
D	time penalty (on same scale as retention time differences, t1 and t2)
timedf	list (length = the number of pairwise alignments) of matrices giving the expected time differences expected at each pair of peaks (used with usePeaks=TRUE.
df	integer, how far from the diagonal to go to calculate the similarity of peaks. Smaller value should run faster, but be careful not to choose too low.
verbose	logical, whether to print out info.
usePeaks	logical, TRUE uses peakdata list, FALSE uses rawdata list for computing similarity.
compress	logical, whether to compress the similarity matrix into a sparse format.

# **Details**

peaksAlignment is a hold-all data structure of the raw and peak detection data.

#### Value

peaksAlignment object

# Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

peaksDataset 25

#### See Also

```
peaksDataset, clusterAlignment
```

#### **Examples**

```
# see clusterAlignment, it calls peaksAlignment
```

peaksDataset Data Structure for raw GCMS data and peak detection results

## **Description**

Store the raw data and optionally, information regarding signal peaks for a number of GCMS runs

## Usage

```
peaksDataset(fns=dir(,"[Cc][Dd][Ff]"), verbose=TRUE, mz=seq(50,550), rtDivide=60, rtrange=NULL)
```

# Arguments

fns character vector, filenames of raw data in CDF format.

verbose logical, if TRUE then iteration progress information is output.

mz vector giving bins of raw data table.

rtDivide number giving the amount to divide the retention times by.

rtrange retention time range to limit data to (must be numeric vector of length 2)

# **Details**

peaksDataset is a hold-all data structure of the raw and peak detection data.

# Value

peaksDataset object

#### Author(s)

Mark Robinson

# References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

26 plot.peaksDataset

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
show(pd)</pre>
```

plot.peaksDataset

Plotting functions for GCMS data objects

#### **Description**

Store the raw data and optionally, information regarding signal peaks for a number of GCMS runs

#### Usage

## **Arguments**

object a peaksDataset, peaksAlignment or clusterAlignment object.

runs for peaksDataset only: set of run indices to plot

mzind for peaksDataset only: set of mass-to-charge indices to sum over (default, all)

mind for peaksDataset only: matrix of aligned indices

plotSampleLabels

for peaksDataset only: logical, whether to display sample labels

calcGlobalMax for peaksDataset only: logical, whether to calculate an overall maximum for

scaling

peakCex character expansion factor for peak labels

plotPeaks for peaksDataset only: logical, whether to plot hashes for each peak

plot.peaksDataset 27

plotPeakBoundaries

for peaksDataset only: logical, whether to display peak boundaries

plotPeakLabels for peaksDataset only: logical, whether to display peak labels

plotMergedPeakLabels

for peaksDataset only: logical, whether to display 'merged' peak labels

mlwd for peaksDataset only: line width of lines indicating the alignment

usePeaks for peaksDataset only: logical, whether to plot alignment of peaks (otherwise,

scans)

plotAcrossRuns for peaksDataset only: logical, whether to plot across peaks when unmatched

peak is given

overlap for peaksDataset only: logical, whether to plot TIC/XICs overlapping

rtrange for peaksDataset only: vector of length 2 giving start and end of the X-axis cols for peaksDataset only: vector of colours (same length as the length of runs) thin for peaksDataset only: when usePeaks=FALSE, plot the alignment lines every

thin values

max.near for peaksDataset only: where to look for maximum

how.near for peaksDataset only: how far away from max.near to look scale.up for peaksDataset only: a constant factor to scale the TICs plotMatches for peaksDataset only: logical, whether to plot matches

xlab for peaksAlignment and clusterAlignment only: x-axis label ylab for peaksAlignment and clusterAlignment only: y-axis label

matchPch for peaksAlignment and clusterAlignment only: match plotting character

matchLwd for peaksAlignment and clusterAlignment only: match line width

matchCex for peaksAlignment and clusterAlignment only: match character expansion

factor

matchCol for peaksAlignment and clusterAlignment only: match colour

for peaksAlignment and clusterAlignment only: vector of colours for colourscale breaks for peaksAlignment and clusterAlignment only: vector of breaks for colourscale alignment for peaksAlignment and clusterAlignment only: the set of alignments to plot

... further arguments passed to the plot or image command

#### **Details**

For peakDataset objects, each TIC is scale to the maximum value (as specified by the how.near and max.near values). The many parameters gives considerable flexibility of how the TICs can be visualized.

For peakAlignment objects, the similarity matrix is plotted and optionally, the set of matching peaks. clusterAlignment objects are just a collection of all pairwise peakAlignment objects.

# Author(s)

Mark Robinson

28 plotImage

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

## See Also

```
plotImage, peaksDataset
```

## **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data
pd<-peaksDataset(cdfFiles[1:3],mz=seq(50,550),rtrange=c(7.5,8.5))

# image plot
plot(pd,rtrange=c(7.5,8.5),plotPeaks=TRUE,plotPeakLabels=TRUE)</pre>
```

plotImage

Plot of images of GCMS data

## **Description**

```
Image plots (i.e. 2D heatmaps) of raw GCMS profile data
```

#### Usage

```
plotImage(object,run=1,rtrange=c(11,13),main=NULL,mzrange=c(50,200),SCALE=log2,...)
```

#### **Arguments**

object	a peaksDataset object
run	index of the run to plot an image for
rtrange	vector of length 2 giving start and end of the X-axis (retention time)
main	main title (auto-constructed if not specified)
mzrange	vector of length 2 giving start and end of the Y-axis (mass-to-charge ratio)
SCALE	function called to scale the data (default: log2)
	further arguments passed to the image command

plotSpectra 29

# **Details**

For peakDataset objects, each TIC is scale to the maximum value (as specified by the how.near and max.near values). The many parameters gives considerable flexibility of how the TICs can be visualized.

For peakAlignment objects, the similarity matrix is plotted and optionally, the set of matching peaks. clusterAlignment objects are just a collection of all pairwise peakAlignment objects.

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
plot, peaksDataset
```

## **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data
pd<-peaksDataset(cdfFiles[1],mz=seq(50,550),rtrange=c(7.5,8.5))

# image plot
plotImage(pd,run=1,rtrange=c(7.5,8.5),main="")</pre>
```

plotSpectra

plotSpectra

#### **Description**

Plot the mass spectra from the profile matrix

#### Usage

```
plotSpectra(object, sample, spectraID, normalize = TRUE, ...)
```

30 plotSpectra

# Arguments

object	an object of class "peaksDataset" where to keep the mass spectra; both abundance (y) than m/ $z\left(x\right)$
sample	character, the sample from were to plot the mass spectra
spectraID	numerical, a vector containing the index of the spectra to be plotted.
normalize	logical, if TRUE normalize the intensity of the mass peak to 100, the most abundant is 100 other peaks are scaled consequetially
	other parameter passed to the plot() function

#### **Details**

Plot the deconvoluted mass spectra from the profile matrix

#### Author(s)

riccardo.romoli@unifi.it

# **Examples**

```
## need access to CDF (raw data)
require(gcspikelite)
gcmsPath <- paste(find.package("gcspikelite"), "data", sep="/")</pre>
## full paths to file names
cdfFiles <- dir(gcmsPath, "CDF", full=TRUE)</pre>
## create a 'peaksDataset' object and add XCMS peaks to it
pd <- peaksDataset(cdfFiles[1:3], \ mz = seq(50,550), \ rtrange = c(7.5,8.5))\\
pd <- addXCMSPeaks(cdfFiles[1:3], pd, peakPicking=c('mF'), snthresh=5,</pre>
                    fwhm=10, step=1, steps=2, mzdiff=0.5, sleep=0)
## align the chromatograms
mp <- correlationAlignment(object=pd, thr=0.8, D=20,</pre>
                            penality=0.2, normalize=TRUE,
                            minFilter=2)
## view the alignment results
mp@Alignment
## plot the mass spectra
par(mfrow=c(3,1))
plotSpectra(object=pd, sample=cdfFiles[1], spectraID=2)
plotSpectra(object=pd, sample=cdfFiles[2], spectraID=3)
plotSpectra(object=pd, sample=cdfFiles[3], spectraID=4)
```

progressiveAlignment-class

Data Structure for progressive alignment of many GCMS samples

# Description

Performs a progressive peak alignment (clustalw style) of multiple GCMS peak lists

#### Usage

progressiveAlignment(pD,cA,D=1000,gap=.5,verbose=TRUE,usePeaks=TRUE,df=30,compress=TRUE)

## **Arguments**

pD a peaksDataset object
cA a clusterAlignment object
D retention time penalty

gap gap parameter

verbose logical, whether to print information

usePeaks logical, whether to use peaks (if TRUE) or the full 2D profile alignment (if FALSE)

df distance from diagonal to calculate similarity

compress logical, whether to store the similarity matrices in sparse form

#### **Details**

The progressive peak alignment we implemented here for multiple GCMS peak lists is analogous to how clustalw takes a set of pairwise sequence alignments and progressively builds a multiple alignment. More details can be found in the reference below.

# Value

progressiveAlignment object

## Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

peaksDataset, multipleAlignment

32 retFatMatrix

#### **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])

ca<-clusterAlignment(pd, gap = .5,D=.05,df=30)
pa<-progressiveAlignment(pd, ca, gap = .6, D=.1,df=30)</pre>
```

retFatMatrix

retFatMatrix

## **Description**

Build a fat data matrix

#### Usage

```
retFatMatrix(data)
```

#### **Arguments**

data

the list obtained from gatherInfo()

#### **Details**

This function allows to extract the data from an object created using gatherInfo and build a data matrix using the area of the deconvoluted and aligned peaks. The row are the samples while the column represent the different peaks.

#### Value

A fat data matrix containing the area of the deconvoluted and aligned peaks. The row are the samples while the column represent the different peaks.

## Author(s)

Riccardo Romoli <riccardo.romoli@unifi.it>

#### See Also

gatherInfo

rmaFitUnit 33

#### **Examples**

rmaFitUnit

Fits a robust linear model (RLM) for one metabolite

## **Description**

Using rlm from MASS, this procedure fits a linear model using all the fragments

## Usage

```
rmaFitUnit(u, maxit=5, mzEffect=TRUE, cls=NULL, fitSample=TRUE, fitOrCoef=c("coef", "fit"), TRANSFORM=log
```

# Arguments

u	a metabolite unit (list object with vectors mz and rt for m/z and retention times, respectively and a data element giving the fragmentxsample intensitity matrix)
maxit	maximum number of iterations (default: 5)
mzEffect	logical, whether to fit m/z effect (default: TRUE)
cls	class variable
fitSample	whether to fit individual samples (alternative is fit by group)
fitOrCoef	whether to return a vector of coefficients (default: "coef"), or an $rlm$ object ("fit")
TRANSFORM	function to transform the raw data to before fitting (default: log2)

#### **Details**

Fits a robust linear model.

34 rmaFitUnit

## Value

list giving elements of fragment and sample coefficients (if fitOrCoef="coef") or a list of elements from the fitting process (if fitOrCoef="fit")

#### Author(s)

Mark Robinson

#### References

Mark D Robinson (2008). Methods for the analysis of gas chromatography - mass spectrometry data *PhD dissertation* University of Melbourne.

#### See Also

```
peaksAlignment, clusterAlignment
```

## **Examples**

```
require(gcspikelite)

# paths and files
gcmsPath<-paste(find.package("gcspikelite"),"data",sep="/")
cdfFiles<-dir(gcmsPath,"CDF",full=TRUE)
eluFiles<-dir(gcmsPath,"ELU",full=TRUE)

# read data, peak detection results
pd<-peaksDataset(cdfFiles[1:2],mz=seq(50,550),rtrange=c(7.5,8.5))
pd<-addAMDISPeaks(pd,eluFiles[1:2])

# pairwise alignment using all scans
fullca<-clusterAlignment(pd, usePeaks = FALSE, df = 100)

# calculate retention time shifts
timedf<-calcTimeDiffs(pd, fullca)</pre>
```

# **Index**

*Topic <b>classes</b>	betweenAlignment-show
betweenAlignment, 6	(betweenAlignment), 6
clusterAlignment, 8	
correlationAlignment, 10	calcTimeDiffs,7
correlationAlignment-class, 12	clusterAlignment, 7, 8, 10, 25, 34
eitherMatrix-class, 14	clusterAlignment-class
multipleAlignment-class, 18	(clusterAlignment), 8
peaksAlignment-class, 24	clusterAlignment-plot
peaksDataset, 25	(clusterAlignment), $8$
plot.peaksDataset, 26	clusterAlignment-show
plotImage, 28	(clusterAlignment), 8
progressiveAlignment-class, 31	compress, 9
*Topic <b>manip</b>	compress,clusterAlignment-method
addAMDISPeaks, 2	(compress), 9
addChromaTOFPeaks, 3	compress,peaksAlignment-method
addXCMSPeaks, 4	(compress), 9
calcTimeDiffs, 7	compress,progressiveAlignment-method
compress, 9	(compress), 9
dp, 13	correlationAlignment, $10, 12$
gatherInfo, 15	correlationAlignment-class, 12
imputePeaks, 17	correlationAlignment-show
normDotProduct, 20	<pre>(correlationAlignment-class),</pre>
parseChromaTOF, 21	12
parseELU, 22	
rmaFitUnit, 33	decompress (compress), 9
.plotcA(plot.peaksDataset), 26	decompress, clusterAlignment-method
.plotpA (plot.peaksDataset), 26	(compress), 9
.plotpD (plot.peaksDataset), 26	decompress, peaksAlignment-method
	(compress), 9
addAMDISPeaks, 2, 22, 23	decompress, progressiveAlignment-method
addChromaTOFPeaks, 3, 22	(compress), 9 dp, 13, 21
addXCMSPeaks, 4, 11	up, 13, 21
align,correlatioAligment	eitherMatrix-class, 14
(correlationAlignment-class),	exportSpectra, 15
12	export espectra, 15
	findPeaks.centWave,5
betweenAlignment, 6, 19	findPeaks.matchedFilter, 5
betweenAlignment-class	,
(betweenAlignment), 6	gatherInfo, 15, 32

36 INDEX

imputePeaks, 16, 17	progressiveAlignment-class, 31
	progressiveAlignment-show
multipleAlignment, 6, 18, 31	<pre>(progressiveAlignment-class),</pre>
multipleAlignment	31
(multipleAlignment-class), 18	
multipleAlignment-class, 18	retFatMatrix, 32
multipleAlignment-show	rmaFitUnit, 33
(multipleAlignment-class), 18	
	show,betweenAlignment-method
normDotProduct, 13, 20	(betweenAlignment), 6
	show,clusterAlignment-method
parseChromaTOF, 4, 21	(clusterAlignment), 8
parseELU, <i>3</i> , 22	show,correlationAlignment-method
peaksAlignment, 7, 9, 10, 14, 21, 34	<pre>(correlationAlignment-class),</pre>
<pre>peaksAlignment (peaksAlignment-class),</pre>	12
24	show,correlationAlignment-method,
peaksAlignment-class, 24	ANY-methods
peaksAlignment-plot	<pre>(correlationAlignment-class),</pre>
(peaksAlignment-class), 24	12
peaksAlignment-show	show, multipleAlignment-method
(peaksAlignment-class), 24	(multipleAlignment-class), 18
peaksDataset, 3-5, 9, 11, 18, 19, 25, 25, 28,	show,peaksAlignment-method
29. 31	(peaksAlignment-class), 24
peaksDataset-class (peaksDataset), 25	show, peaksDataset-method
peaksDataset-plot (peaksDataset), 25	(peaksDataset), 25
peaksDataset-show (peaksDataset), 25	show,progressiveAlignment-method
plot, 29	(progressiveAlignment-class),
plot (plot.peaksDataset), 26	31
plot,clusterAlignment,ANY-method	
(clusterAlignment), 8	
plot,clusterAlignment-method	
(clusterAlignment), 8	
plot, peaksAlignment, ANY-method	
(peaksAlignment-class), 24	
plot, peaksAlignment-method	
(peaksAlignment-class), 24	
plot, peaksDataset, ANY-method	
(peaksDataset), 25	
plot,peaksDataset-method	
(peaksDataset), 25	
plot.peaksDataset, 26	
plotImage, 28, 28	
plotImage,peaksDataset-method	
(plotImage), 28	
plotSpectra, 29	
progressiveAlignment, 10, 18, 19	
progressiveAlignment	
<pre>(progressiveAlignment-class),</pre>	
31	