Peak Gather Example

bjp

6/29/2021

Example of gathering data for DB input

Include all necessary packages:

```
source('src/base/mzML_base.R')
```

```
## Loading required package: XML
```

```
## Loading required package: base64enc
```

```
source('src/methodreportingtool/parse_methodxml.R')
source('src/gather/table_msdata.R')
source('src/gather/peak_gather.R')
source('src/base/peaktable2.R')
```

Step 1) Load Method XML data and mzML data

```
xmlfile <- 'example/PFAC30PAR_PFCA2_mzML.XML'
methodxml <- parse_methodxml(xmlfile)
mzmlfile <- paste(dirname(xmlfile), "/", methodxml$method$sample$filename, sep = "")
mzml <- mzMLtoR(mzmlfile)
```

XML Data

Sample Info:

filename	PFAC30PAR_PFCA2.mzML
description	Reference Standard for PFAS
class	analytical standard
submitter	bjp@nist.gov

Chromatography Info:

ctype	Liquid Chromatography
cvendor	ThermoFisher Scientific
cmodel	UltiMate 3000

ssolvent	water
mp1solvent	water
mp1add	ammonium acetate
m2solvent	methanol
mp2add	ammonium acetate
mp3solvent	none
mp3add	none
mp4solvent	none
mp4add	none
colvendor	Agilent Technologies
colname	Poroshell C18
colchemistry	C18
colid	2.1
collen	50
coldp	2.7
gcolvendor	none
gcolname	NULL
gcolchemistry	none
gcolid	NULL
gcollen	NULL
gcoldp	NULL

Mass Spectrometry Info:

msvendor	ThermoFisher Scientific
msmodel	Q-Exactive
imode	electrospray ionization
polarity	negative
vvalue	2500
vunits	V
massanalyzer1	quadrupole
massanalyzer2	orbitrap
fragmode	HCD
cevalue	30
cetype	fixed
ceunits	normalized
ms2exp	DDA
isowidth	0.7
msaccuracy	5
ms1resolution	70000
ms2resolution	17500

QC Method Info:

qcused	True
qctype	True
qctype	True
qctype	False
qctype	False

Peaklist:

name	identifier	ionstate	mz	rt	peak_starttime	peak_endtime	verified	.attrs
Perfluoropentanoic acid	2646	[M-H]-	262.976	8.6	8.4	9.1	True	1
Perfluorohexanoic acid	2643	[M-H]-	312.973	10.8	10.5	11.1	True	2
Perfluoroheptanoic acid	2640	[M-H]-	362.9699	12.09	11.9	12.4	True	3
Perfluorooctanoic acid	2637	[M-H]-	412.9665	13.05	12.8	13.4	True	4
Perfluorononanoic acid	2635	[M-H]-	462.9635	13.8	13.6	14.1	True	5
Perfluorodecanoic acid	2632	[M-H]-	512.9602	14.5	14.3	14.8	True	6
Perfluoroundecanoic acid	2630	[M-H]-	562.9573	15.1	14.9	15.4	True	7
Perfluorododecanoic acid	2629	[M-H]-	612.954	15.6	15.4	15.8	True	8
Perfluorotridecanoic acid	2628	[M-H]-	662.9516	16	15.9	16.2	True	9
Perfluorotetradecanoic acid	2627	[M-H]-	712.9487	16.3	16.2	16.6	True	10

Step 2) For each peak in the peak list, gather method data, peak data, and peak-specific MS data.

```
dat <- peak_gather(methodxml, mzml)
```

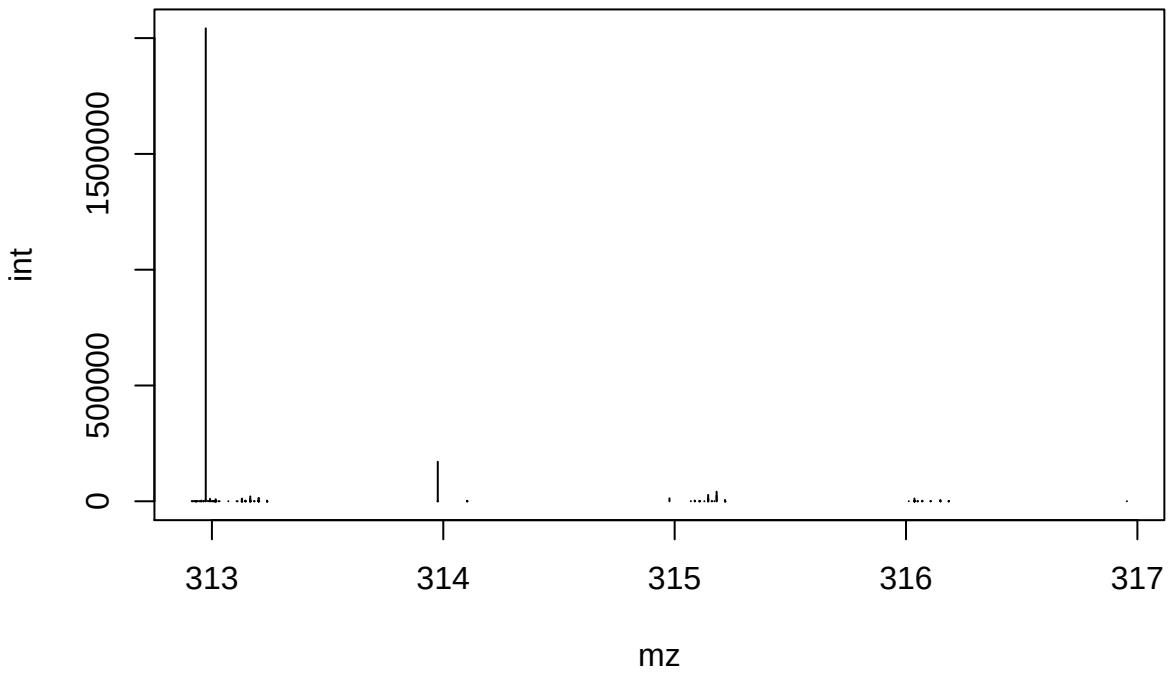
Example dataset:

```
i <- 2 ## because there were no annotations for peak 1
dat[[i]]$peak
```

```
##                         name identifier ionstate      mz   rt peak_starttime
## count Perfluorohexanoic acid       2643   [M-H]- 312.973 10.8           10.5
##       peak_endtime verified .attrs
## count         11.1     True      2
```

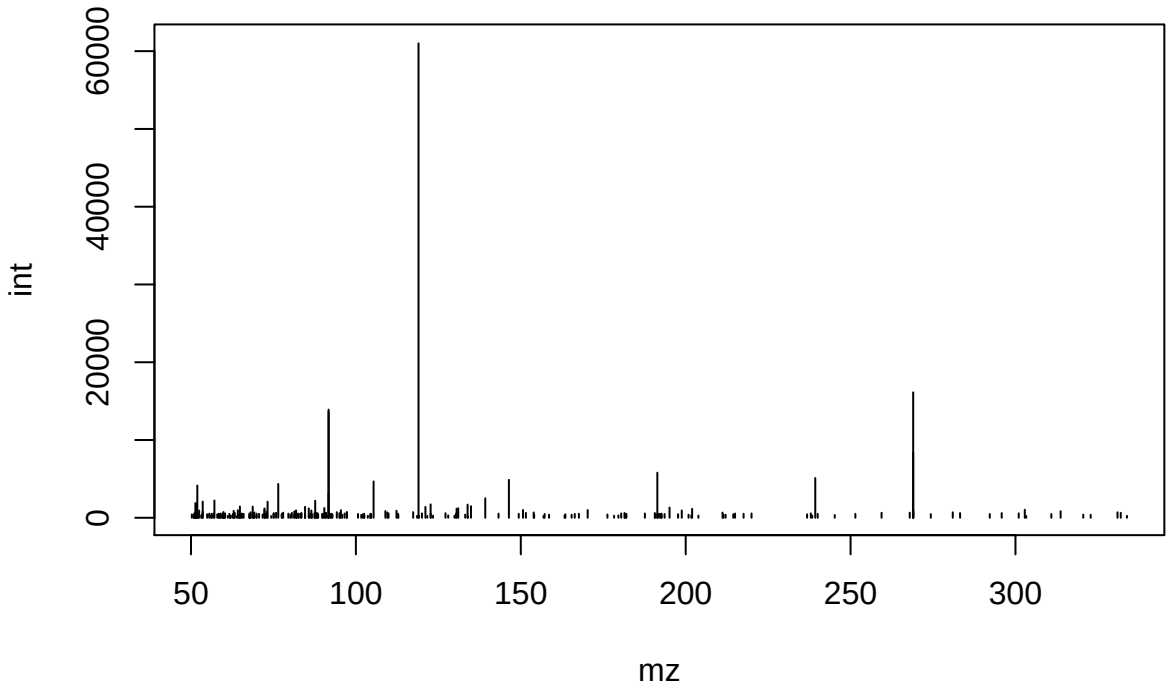
MS1

```
ms1_peaktable <- peaktable(dat[[i]]$ms1data$msdata, masserror =  as.numeric(dat[[i]]$method$massspecti
mz <- rowMeans(ms1_peaktable$mass, na.rm = TRUE)
int <- rowMeans(ms1_peaktable$int, na.rm = TRUE)
ind <- which(mz >= as.numeric(dat[[i]]$peak$mz) - 0.5 & mz <= as.numeric(dat[[i]]$peak$mz) + 4)
mz <- mz[ind]
int <- int[ind]
plot(mz, int, type = "h")
```



MS2

```
ms2_peaktable <- peaktable(dat[[i]]$ms2data$msdata, masserror =  as.numeric(dat[[i]]$method$massspecti
mz <- rowMeans(ms2_peaktable$mass, na.rm = TRUE)
int <- rowMeans(ms2_peaktable$int, na.rm = TRUE)
plot(mz, int, type = "h")
```



Annotations:

```
knitr::kable(dat[[i]]$annotation)
```

fragment_mz	fragment_formula	fragment_SMILES	fragment_radical	fragment_citation
118.9912	C2F5	FC-C(F)(F)F	False	DOI:10.1002/rcm.3274
168.9883	C3F7	FC-C(F)(F)C(F)(F)F	False	DOI:10.1002/rcm.3274
318.9796	C6F13	FC(F)(C(F)(F)C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)F	False	DOI:10.1002/rcm.3274

write file for future reference

```
saveRDS(dat, 'example/PFAC30PAR_PFCA2_output.RDS')
```