Aggregate functions (Example)

1) Install Dependencies and Initialize Environment

```
#install dependencies and verify directories
source('src/compliance.R')

#copy example files to Input folder
examples <- list.files(path = "example", full.names = TRUE)
file.copy(examples, 'input', overwrite = TRUE)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE
```

Within the environment, the object *'fragments'* was included from *'src/aggregate/fragment_table.csv'* containing a list of possible fragments with fragment-specific information available.

FRAGMENT_ID	FRAGMENT_FORMULA	FRAGMENT_EXACTMASS	FRAGMENT_DESCRIPTION
1	CF3	68.9958	trifluoromethanide
2	SO3	79.9579	sulfonate (oxido(dioxo)-?6-sulfanyl) rad
3	SO3F	98.9558	sulfurofluoridate
4	C2F5	118.9926	1,1,2,2,2-pentafluoroethan-1-ide
5	CF3SO3	148.9531	difluoromethanesulfonate
6	C3F7	168.9894	1,1,2,2,3,3,3-heptafluoropropan-1-ide

2) Import Peak List

A peak list consists of the minimum information needed to extract compound-specific MS data from the designated mzML or other raw data file. Metadata files, such as the LC Method (if separate from the MS Method), MS Method, and QC Method are included for aggregation.

```
peaklist_loc = "input/example_peaklist.csv"

peaklist <- read.csv(peaklist_loc, header = TRUE, row.names = NULL)

knitr::kable(peaklist, format = 'latex')
```

ID	FILENAME	SAMPLE_TYPE	DB_IDENTIFIER	M.Z	RT
506	PFAC30PAR_PFCA2.mzML	Standard	YPJUNDFVDDCYIH-UHFFFAOYSA-N	212.9792	4.50
507	PFAC30PAR_PFCA2.mzML	Standard	CXZGQIAOTKWCDB-UHFFFAOYSA-N	262.9760	8.60
508	PFAC30PAR_PFCA2.mzML	Standard	PXUULQAPEKKVAH-UHFFFAOYSA-N	312.9728	10.80
509	PFAC30PAR_PFCA2.mzML	Standard	ZWBAMYVPMDSJGQ-UHFFFAOYSA-N	362.9696	12.09
510	PFAC30PAR_PFCA2.mzML	Standard	SNGREZUHAYWORS-UHFFFAOYSA-N	412.9664	13.05
511	PFAC30PAR_PFCA2.mzML	Standard	UZUFPBIDKMEQEQ-UHFFFAOYSA-N	462.9632	13.80
512	PFAC30PAR_PFCA2.mzML	Standard	PCIUEQPBYFRTEM-UHFFFAOYSA-N	512.9600	14.50
513	PFAC30PAR_PFCA2.mzML	Standard	CXGONMQFMIYUJR-UHFFFAOYSA-N	612.9537	15.60
514	PFAC30PAR_PFCA2.mzML	Standard	SIDINRCMMRKXGQ-UHFFFAOYSA-N	562.9568	15.10
515	PFAC30PAR_PFCA2.mzML	Standard	LVDGGZAZAYHXEY-UHFFFAOYSA-N	662.9505	16.00
516	PFAC30PAR_PFCA2.mzML	Standard	RUDINRUXCKIXAJ-UHFFFAOYSA-N	712.9473	16.40

3) Search fragments & aggregate peaklist data

The *aggregate_fragments* function extracts MS2 data for each compound, based on the peak list information, and attempts to identify any annotatable fragments from the *fragments* object, based on instrument accuracy.

```
masserror = 10 #instrument accuracy (ppm)
minerror = 0.002 #minimum instrument accuracy (Da)
correl = NULL #correlation coefficient for MS2 ions to MS1 precursor ion
ph = NULL #minimum chromatographic peak height (%) above which to include MS2 ions
freq = NULL #minimum observational frequency (%) for MS2 ions to be included
# For more information on these parameters, see DOI: 10.1021/jasms.0c00423

peaklist_aggregate <- aggregate_fragments(peaklist, fragments, masserror, minerror, correl, ph, freq)

saveRDS(peaklist_aggregate, file = 'output/example/peaklist_aggregate_example.RDS')
```

Based on the example aggregation, there are 10 compounds with annotatable MS2 fragments.