LC-MS Peak Annotation and Identification with *CAMERA*

Carsten Kuhl, Ralf Tautenhahn and Steffen Neumann

April 21, 2009

Introduction

The R-package *CAMERA* is a (**C**ollection of **A**lgorithms for **ME**tabolite p**R**ofile **A**nnotation). Primarily used to annotate LC-MS data. It is designed to interact directly with processed data from *xcms* and additional analyses with *Rdisop*.
It includes the annotation of isotope peaks, adducts and fragments in peak lists generated by *xcms*. A set of annotation methods is used, which group together mass signals measured from a single metabolite, based on rules for mass differences and peak shape comparison [1].
Based on this annotation, the molecular composition can be calculated if the mass spectrometer has a high-enough accuracy for both the mass and the isotope pattern intensities.

1 Peak Annotation

1.1 Adduct list and molecular mass estimation

For soft ionisation methods such as LC/ESI-MS, different adducts (e.g. $[M+K]^+$, $[M+Na]^+$) and fragments (e.g. $[M-C_3H_9N]^+$, $[M+H-H_20]^+$) occur. Depending on the molecule having an intrinsic charge, $[M]^+$ may be observed as well. An estimation of the molecular mass of $[M]$ can be calculated from at least two annotated adduct ions. To scan for adducts every theoretical possible combination of adducts from a given list of ions are calculated. For a small example see Tab. 1. Every group of peaks is scanned, if these combinations fits with the mass differences and then molecular masses are computated.

Formula	Mass difference in amu
$[M+H]^+$	1.007276
$[M+Na]^+$	22.98977
$[M+K]^+$	38.963708
$[2M+Na]^+$	22.98977
$[M+H+Na]^{2+}$	23.9976
...	...

Table 1: Examples of calculated adducts for the Kations (K,H,Na) with their mass differences occuring in positive ion mode. The actual difference is calculated considering the charge and the number of molecules M in the observed ion.

2 Processing with *CAMERA*

2.1 Preprocessing with *xcms*

At first, create an *xcmsSet* with your favourite parameters, e.g.:

```
library(CAMERA)
file <- system.file('mzdata/MM14.mzdata', package = "CAMERA")
xs <- xcmsSet(file,method="centWave",ppm=30,peakwidth=c(5,10))
```

2.2 Annotation

The annotatation wants to answer the questions which peaks belongs together and computate the exact mass of the molecule, from which the ions originate. A annotation for an *xcmsSet* with one sample can be done quick with

```
library(CAMERA)
an <- annotate(xs)
peaklist <- getPeaklist(an)
```

Note: if xs is a grouped xcmsSet, CAMERA needs additional parameters
The annotation workflow contains following steps:

1. peak grouping after retention time
2. peak group verification with EICs correlation
3. annotation of possible isotopes
4. annotation of adducts and calculating hypothetical masses for the group

The result of an annotaton an is a data-frame similiar to a peak table and can be easily stored in a comma separated table (Excel-readable).

```
write.csv(peaklist,file='xsannotated.csv')
```

2.2.1 Paramater for `annotate`

Annotate is a wrapper function for many *CAMERA* S4 methods, so every parameter of the methods can be transfered via annotate. The parameters of these functions are here short summerised. For addional information see the manpage of `annotate`.
For the peak grouping a retention time window is calculated after the FWHM (full width half maximum) of the local highest peak. Therefore one additional parameter can be pass for the FWHM calculation (*perfwhm = 0.3*, which is 30% of used FWHM width).
The peak group verification step use a pearson correlation with the parameter *cor_eic_th = 0.75*, which is the height of the peak correlation, that two peak must have to be consider to originate from the same molecule.
The annotation of isotopes and adducts share the parameters (*ppm = 5*) and (*mzabs = 0.01*), which are the relative and absolute error for m/z diffs.
Another parameters for isotope finding are the maximum charge (*maxcharge = 3*) and maximum number of isotopes (*maxiso = 4*) which are expectate to occur. For addional information see `findIsotopes`.
The adduct annotation has one additional parameter (*multiplier = 3*), which is the maximum number n of molecules in clusterions (e.g. [nM+H]). For information about creating a ruleset fÃijr annotation see `findAdducts`.
If the xcmsSet contains more than one sample or several different classes e.g. "wildtype" and "knockout", you must choose which one should be annotated. Therefore the parameter sample and category exist. For an example see section 3.
So with more parameters a call of annotate looks like:

```
an <- annotate(xs,sigma=6, perfwhm=0.3, cor_eic_th=0.75,
maxcharge=3, maxiso=3, ppm=5, mzabs=0.01,polarity="positive")
```

2.3 Annotation without verification by correlation

A short notice for former *esi* user, this step is now obsolete and not longer supported. All annotations use the peak correlation if possible.

2.4 Interpretation of the Results

Table 2 shows an example of annotation results. A small cutout of the result table is displayed, the columns with the intensity values are omitted and the rows are ordered by there rt values for better readability. The column *pc* shows the result of the peak correlation based annotation (independent of the annotations *iso* and *adduct*). Peaks with the same label are supposed to

id	mz	rt	isotopes	adduct	pc
65	176.04	280.09			
76	136.05	280.43	[14][M+1]1+		5
77	135.05	280.43	[14][M]1+		5
74	153.06	280.43		[M+H]+ 152.05437	5
75	175.04	280.43		[M+Na]+ 152.05437	5
73	197.02	280.76		[M+2Na-H]+ 152.05437	5
78	377.74	286.15			
79	732.5	286.49			
83	488.32	286.82		[M+Na]+ 465.33205	7
82	466.34	286.82		[M+H]+ 465.33205	7
...					

Table 2: Example of annotation results. Colums with intensity values are omitted. blue-line: annotated group 5, red-line: annotated group 7

belong to the same spectrum. The column *adduct* shows the annotation hypotheses for the ions. The value after the brackets is the estimated molecular mass.
The column *isotopes* contains the annotated isotopes for a monoisotopic peak. The values in the first square brackets denote the isotope-group-id(column *id*), the second is the isotope annotation and the number after the brackets is the charge of the isotope.

3 Examples using CAMERA test dataset

Example 1 Fast annotation without further using of xsAnnotate of the MM14 dataset.

```
library(CAMERA)
file <- system.file('mzdata/MM14.mzdata', package = "CAMERA")
xs <- xcmsSet(file,method="centWave",ppm=30,peakwidth=c(5,10))
an <- annotate(xs)
peaklist <- getPeaklist(an)
write.csv(peaklist,'/tmp/mm14.csv')
```

There are 126 peaks in 10 groups, of which 48 peaks get isotope annotations and 25 peaks are annotated as adducts. length(which(peaklist[,"adduct"]!=""))

Example 2 Annotation with exact use of an xsAnnotate object.

```
library(CAMERA)
cdfpath <- system.file("cdf", package = "faahKO")
cdffiles <- list.files(cdfpath, recursive = TRUE,full=T)

xset <- xcmsSet(cdffiles,snthresh=3,max=10)
xsg <- group(xset)
xsg <- retcor(xsg)
xsg <- group(xsg,bw=10)

#create xsAnnotate object
xanno<-xsAnnotate(xsg,sample=1,category="WT")

#group according to retention time
xanno<-groupFWHM(xanno)

#check grouping with EIC correlation, when indicated regroup
xanno<-groupCorr(xanno)

#search for isotopes
xanno<-findIsotopes(xanno)

#calculate possible adducts
xanno<-findAdducts(xanno,polarity="positive")

#get annotated peaklist
an<-getPeaklist(xanno)
write.csv(an,'/tmp/faah-an2.csv')
```

There are 1829 peaks in 221 groups, of which 126 peaks get isotope annotations and 126 peaks are annotated as adducts.

References

[1] Ralf Tautenhahn, Christoph Böttcher, Steffen Neumann : Annotation of LC/ESI–MS Mass Signals, BIRD 2007 Proc. of BIRD 2007 – 1st International Conference on Bioinformatics Research and Development, 2007. http://www.springerlink.com/content/4731404001787974/ and http://msbi.ipb-halle.de/~rtautenh/bird07.pdf