

# Isotope cluster detection and validation

Hendrik Treutler

October 27, 2020

## 1 Introduction

In [1] we propose an approach for the exhaustive detection and mass-specific validation of isotope clusters. In this approach we perform an targeted peak picking for the detection of isotopologue features and we perform a mass-specific validation of putative isotope clusters. The approach is implemented in the following workflow using functions from *xcms* [2] and *CAMERA* [3]. First, we perform peak picking without any prior knowledge which we denote as *traditional peak picking*. Here, we use the *centWave* algorithm [4] which applies a heuristic for the detection of ROIs (ROI identification step). Given the set of detected ROIs, chromatographic peaks are extracted using continuous wavelet transformation (ROI examination step). This step results in a peak table with one row for each detected feature and one column for each feature property such as  $m/z$ , retention time, integrated peak area, and signal-to-noise ratio.

Second, we perform the proposed targeted peak picking as described in [1]. Here, a set of isotope ROIs is predicted on basis of the previously extracted peak table. Given the set of predicted isotope ROIs, chromatographic peaks are extracted using continuous wavelet transformation (ROI examination step). Notably, this ROI examination step is identical to the ROI examination step in the traditional peak picking step with the exception that we use relaxed peak picking parameters this time. This step results in an extended peak table which is enriched with features corresponding to isotope peaks.

Third, we extract *pseudospectra* from the extended peak table [3]. This step aims at the extraction of compound spectra on basis of the retention times, but multiple coeluting compounds are potentially assigned to the same spectrum which is the reason for the usage of the term pseudospectrum. In case of multiple raw data files a retention time correction (*xcms* function *retcor*) can be advisable prior to the extraction of pseudospectra. This step results

in a set of pseudospectra. Each pseudospectrum is a peak table comprising all properties of a subset of the features from the extended peak table.

Fourth, we detect isotope clusters in each pseudospectrum using the proposed isotope detection routine with mass-specific isotope cluster validation as described in [1]. Here, putative isotope clusters are detected and putative isotope clusters are validated based on database statistics. This step results in a set of annotated pseudospectra, i.e. the given set of pseudospectra enriched with isotope annotations.

## 2 Usage

```
## libraries
library("msdata")
library("xcms")
library("CAMERA")

## raw data file
mzdatapath <- file.path(find.package("msdata"), "microtofq")
file <- list.files(path = mzdatapath, pattern = "^.*\\.mzdata$",
                  all.files = TRUE, full.names = TRUE,
                  recursive = FALSE, ignore.case = FALSE,
                  include.dirs = FALSE, no.. = TRUE)[[1]]

## peak picking using centWave in combination with an
## additional peak picking step targeting on isotopes
xs <- xcmsSet(
  files=file, method="centWaveWithPredictedIsotopeROIs",
  ppm=10, peakwidth=c(5,12), snthresh=25, prefilter=c(3,200)
)

## create pseudospectra and perform isotope cluster detection
an <- xsAnnotate(xs, polarity="positive") # constructor; extracts peak table
an <- groupFWHM(an, perfwHM = 1)         # group peaks by retention time
an <- findIsotopesWithValidation(object = an, ppm = 10,
                               mzabs = 0.01, intval="intb",
                               maxcharge = 3) # annotate isotopic peaks

## extract annotated peak table
peakTable <- getPeaklist(an) # extract peak list
```

In the above example, we create an *xcmsSet* object with example data using traditional peak picking in combination with the proposed targeted peak

picking algorithm.

Next, we create an *xsAnnotate* object and create pseudospectra. On basis of the pseudospectra, we perform the proposed isotope detection routine with mass-specific isotope cluster validation.

Finally, an annotated peak table is extracted.

[1] Submitted to Metabolites journal, special issue “Bioinformatics and Data Analysis”.

[2] Smith, C.A.; Want, E.J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* 2006, 78, 779–787.

[3] Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T.R.; Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical chemistry* 2012, 84, 283–289.

[4] Tautenhahn, R.; Böttcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 2008, 9, 504+.