Vignette for SpecLibExample: OrgMassSpecR Mass Spectral Libraries

Nathan G. Dodder

October 2011

Contents

1 Introduction **1**

2 Generating Reports **2**

3 Creating Spectral Library Packages **2**

3.1 Preparing Spectral Files 4

3.2 Preparing Chemical Structure Images 4

3.3 Preparing Metadata 5

1 Introduction

OrgMassSpecR has companion packages that contain mass spectral libraries and functions to compile the libraries into PDF reports[1]. The spectral library packages are available through the OrgMassSpecR R-Forge repository.

Each spectral library package is capable of generating a PDF report. This is intended as the primary mechanism for viewing a library, and provides a way to share a library independent of R. Pre-generated reports are not included in the packages. Instead, links to the reports can be found at OrgMassSpecR project website. A NIST MSP text file can also be generated. This format can be used to import the spectra into the NIST MS Search software as a custom searchable library.

Within the package source, the spectra and metadata are stored in CSV format, and the chemical structures in PNG format. These can be manipulated within R, or copied and used in other software. The spectra consist of centroid m/z values and intensities.

[1] "Library" refers to an organized collection of mass spectra, not a collection of R packages.

The `SpecLibExample` package contains a set exogenous compounds identified in a sea lion blubber sample and is used here to illustrate the file structure and functionality of spectral library packages.

2 Generating Reports

Each package has its own `LibraryReport` function that generates the PDFs. The function is similar across packages; differences exist to accommodate the the specific needs of the library/project. If multiple spectral library packages are loaded, the appropriate function should be called using *`package-name`*`::LibraryReport()`. Each page of the PDF report is one spectrum and the associated metadata (compound name, etc.).

The following example shows how to select one spectrum from the example library and generate a PDF report. Figure 1 shows the second page of the report. The title page is excluded. Alternatively, the spcetrum could be displayed in the R graphic device by setting the `pdf` argument to `FALSE`.

```
> library(SpecLibExample)

> mirex.meta <- example.meta[example.meta$
+   filename == "mirex", ]
> LibraryReport(metadata = mirex.meta, pdf = TRUE)
```

The m/z labels in the mass spectrum refer to the most abundant peak in the cluster. The peak m/z values listed in the table refer to the monoisotopic peak in the cluster.

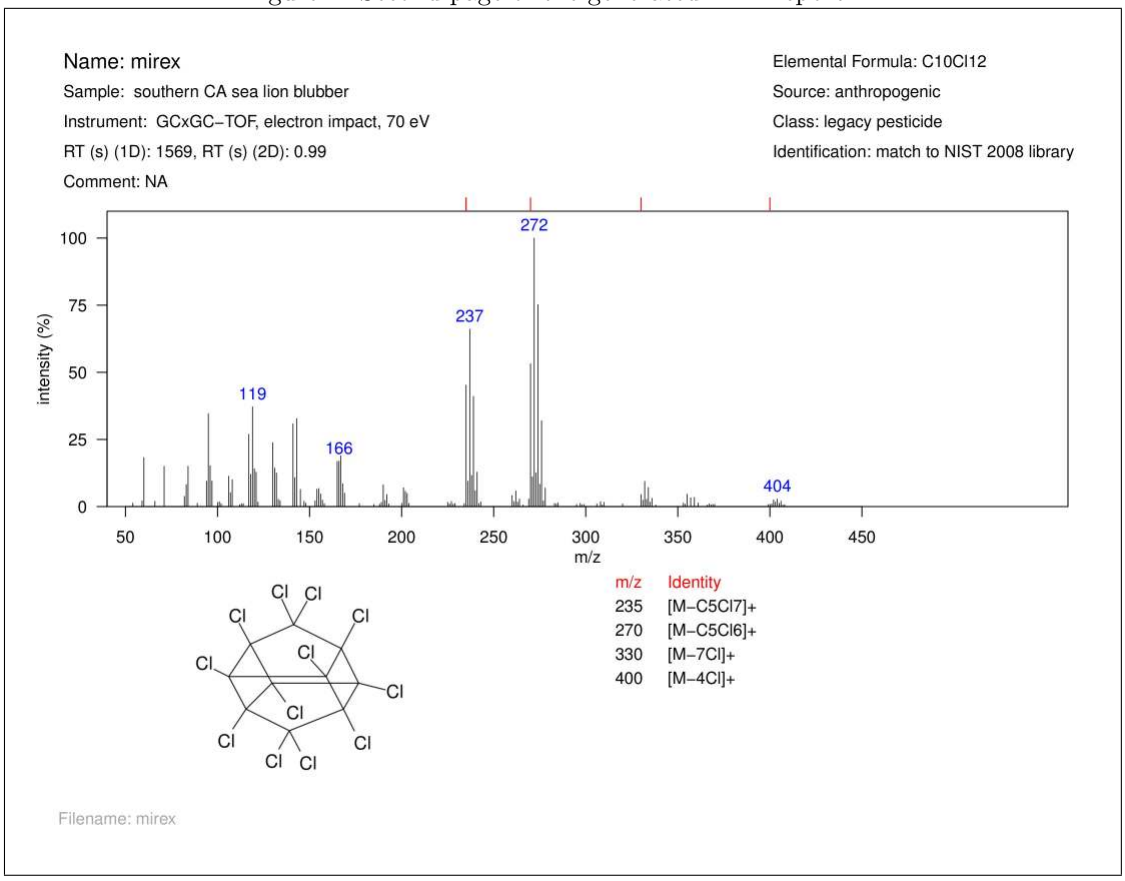
R generates uncompressed PDFs, and the PDFs generated by the OrgMassSpecR spectral library packages can be large (>10 MB). In our experience PDF compression can significantly reduce the file size. This can be performed by Adobe Reader, or see `tools::compactPDF`.

3 Creating Spectral Library Packages

The following are some notes on creating spectral library packages. It is recommended to follow the directory structure and data formats in the `SpecLibExample` package. See the *Writing R Extensions* manual for an introduction to making packages.

The `data` directory contains two CSV files. All mass spectra are stored in a file ending in `.spec.csv`, and the metadata is stored in a file ending in `.meta.csv`. The help documentation for `LibraryReport` describes the required column names. The directory at `/inst/extdata/struct` contains PNG files of the chemical structures.

Figure 1: Second page of the generated PDF report.



`LibraryReport` uses the `filename` defined in the metadata to select the corresponding spectrum and structure, these should therefore be identical (the order does not have to be the same). The row order within `metadata` file controls the page order of the report.

Creating a new package requires modification of the `LibraryReport` function. The title page information is set within the function code.

3.1 Preparing Spectral Files

Spectra for the example library were exported from the instrument's data analysis software as NIST MSP text files. Each MSP file was assigned a unique name. The function `OrgMassSpecR::ReadMspDirectory` was used to assemble the spectra into a single data frame with the format described in the help for `LibraryReport`.

The data frame containing the sea lion spectra is shown below.

```
> head(example.spec)
```

```
                        compound filename mz intensity display identity
1 bis(chlorophenyl)ethane (DDEt)     ddet 50       180   FALSE     <NA>
2 bis(chlorophenyl)ethane (DDEt)     ddet 51       215   FALSE     <NA>
3 bis(chlorophenyl)ethane (DDEt)     ddet 52        54   FALSE     <NA>
4 bis(chlorophenyl)ethane (DDEt)     ddet 62        44   FALSE     <NA>
5 bis(chlorophenyl)ethane (DDEt)     ddet 63       245   FALSE     <NA>
6 bis(chlorophenyl)ethane (DDEt)     ddet 68        19   FALSE     <NA>
```

The `filename` is assigned by `ReadMspDirectory` based on the name of the MSP file. The `display` (`TRUE` or `FALSE`) column specifies if the m/z value is displayed above the corresponding fragment ion peak (usually the most abundant peak in the isotopic cluster). The `identity` (a character string) column specifies the identity of the fragment ion to be printed in the table below the spectrum (usually corresponding to the monoisotopic m/z. We suggest the `display` and `identity` columns be added using a spreadsheet program during the manual identification of the fragment ions. The `compound` column is not strictly required - it was added for clairity.

3.2 Preparing Chemical Structure Images

The PNG files at `/inst/extdata/struct` are read and converted to a raster object using `png::readPNG`. For the example library, the structures were drawn using BKchem, with a font size of 14 and line width of 1 px. They were exported as PNG (Cairo) files at 300 dpi resolution and with a transparent background.

3.3 Preparing Metadata

The required columns for the metadata table are described in the the help for `LibraryReport`. The table was created in a spreadsheet program during the manual identification of the spectra.