

Package: chromConverter (via r-universe)

November 4, 2024

Title Chromatographic File Converter

Version 0.7.0

Maintainer Ethan Bass <ethanbass@gmail.com>

Description Reads chromatograms from binary formats into R objects. Currently supports conversion of 'Agilent ChemStation', 'Agilent MassHunter', 'Shimadzu LabSolutions', 'ThermoRaw', and 'Varian Workstation' files as well as various text-based formats. In addition to its internal parsers, chromConverter contains bindings to parsers in external libraries, such as 'Aston' <<https://github.com/bovee/aston>>, 'Entab' <<https://github.com/bovee/entab>>, 'rainbow' <<https://rainbow-api.readthedocs.io/>>, and 'ThermoRawFileParser' <<https://github.com/compomics/ThermoRawFileParser>>.

License GPL (>= 3)

URL <https://ethanbass.github.io/chromConverter>,
<https://github.com/ethanbass/chromConverter>

BugReports <https://github.com/ethanbass/chromConverter/issues>

Imports bitops, fs, purrr, readxl, reticulate, stringr, tidyr, utils, RaMS, tibble, xml2, bit64, data.table, base64enc, jsonlite, digest

Suggests entab, ncd4, pbapply, testthat (>= 3.0.0), mzR, chromConverterExtraTests

Config/reticulate list(packages = list(list(package = ``scipy"), list(package=``numpy"), list(package = ``pandas"), list(package=``Aston", pip = TRUE), list(package=``rainbow-api", pip = TRUE), list(package = ``olefile", pip = TRUE)))

Encoding UTF-8

Language en-US

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Additional_repositories <https://ethanbass.github.io/drat/>,
<https://ethanbass.r-universe.dev/>

Config/testthat/edition 3

Config/Needs/website rmarkdown, ggplot2, dplyr

Repository <https://ethanbass.r-universe.dev>

RemoteUrl <https://github.com/ethanbass/chromConverter>

RemoteRef HEAD

RemoteSha 767f8671016886a589e6042f50f1f24946c7246c

Contents

call_entab	3
call_openchrom	3
call_rainbow	5
configure_openchrom	6
extract_metadata	7
read_agilent_d	7
read_agilent_dx	8
read_asm	9
read_cdf	10
read_chemstation_ch	11
read_chemstation_csv	12
read_chemstation_ms	13
read_chemstation_reports	14
read_chemstation_uv	15
read_chromeleon	16
read_chroms	17
read_mdf	19
read_mzml	20
read_peaklist	21
read_shimadzu	22
read_shimadzu_gcd	23
read_shimadzu_lcd	24
read_shimadzu_qgd	26
read_sz_lcd_2d	28
read_sz_lcd_3d	29
read_thermoraw	30
read_varian_peaklist	32
read_varian_sms	32
read_waters_arw	34
read_waters_raw	35
sp_converter	36
uv_converter	36
write_andi_chrom	37
write_chroms	38
write_mzml	39

call_entab	<i>Call Entab Converts files using Entab parsers</i>
------------	--

Description

Call Entab Converts files using Entab parsers

Usage

```
call_entab(
  path,
  data_format = c("wide", "long"),
  format_out = c("matrix", "data.frame", "data.table"),
  format_in = "",
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw")
)
```

Arguments

path	Path to file
data_format	Whether to return data in wide or long format.
format_out	Class of output. Either matrix, data.frame, or data.table.
format_in	Format of input.
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either chromconverter or raw.

Value

A chromatogram in the format specified by format_out (retention time x wavelength).

call_openchrom	<i>Parse files with OpenChrom</i>
----------------	-----------------------------------

Description

Writes xml batch-files and calls OpenChrom file parsers using a system call to the command-line interface. To use this function **OpenChrom** (version 0.4.0) must be manually installed. The command line interface is no longer supported in the latest versions of OpenChrom (starting with version 0.5.0), so the function will no longer work with these newer versions. OpenChrom 1.4 has been scrubbed from the internet, but OpenChrom 1.3 is still available from [SourceForge](#).

Usage

```
call_openchrom(  
  files,  
  path_out = NULL,  
  format_in,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  export_format = c("mzml", "csv", "cdf", "animl"),  
  return_paths = FALSE,  
  verbose = getOption("verbose")  
)
```

Arguments

files	files to parse
path_out	directory to export converted files.
format_in	Either msd for mass spectrometry data, csd for flame ionization data, or wsd for DAD/UV data.
format_out	R format. Either matrix, data.frame or data.table.
data_format	Whether to return data in wide or long format.
export_format	Either mzml, csv, cdf, animl. Defaults to mzml.
return_paths	Logical. If TRUE, the function will return a character vector of paths to the newly created files.
verbose	Logical. Whether to print output from OpenChrom to the console.

Details

The `call_openchrom` works by creating an xml batchfile and feeding it to the OpenChrom command-line interface. OpenChrom batchfiles consist of `InputEntries` (the files you want to convert) and `ProcessEntries` (what you want to do to the files). The parsers are organized into broad categories by detector-type and output format. The detector-types are `msd` (mass selective detectors), `csd` (current selective detectors, e.g. FID, ECD, NPD), and `wsd` (wavelength selective detectors, e.g. DAD, and UV/VIS). Thus, when calling the OpenChrom parsers, you must select one of these three options for the input format (`format_in`).

Value

If `return_paths` is FALSE, the function will return a list of chromatograms (if an appropriate parser is available to import the files into R). The chromatograms will be returned in `matrix` or `data.frame` format according to the value of `format_out`. If `return_paths` is TRUE, the function will return a character vector of paths to the newly created files.

Side effects

Chromatograms will be exported in the format specified by `export_format` in the folder specified by `path_out`.

Note

Activating the OpenChrom command-line will deactivate the graphical user interface (GUI). Thus, if you wish to continue using the OpenChrom GUI, it is recommended to create a separate command-line version of OpenChrom to call from R.

Author(s)

Ethan Bass

References

Wenig, Philip and Odermatt, Juergen. OpenChrom: A Cross-Platform Open Source Software for the Mass Spectrometric Analysis of Chromatographic Data. *BMC Bioinformatics* **11**, no. 1 (July 30, 2010): 405. doi:10.1186/1471210511405.

call_rainbow

Parse Agilent or Waters files with rainbow parser

Description

Uses **rainbow** parsers to read in Agilent (.D) and Waters (.raw) files. If format_in is agilent_d or waters_raw, a directory of the appropriate format (.d or .raw) should be provided to file. If format_in is chemstation_uv a .uv file should be provided. Data can be filtered by detector type using the what argument.

Usage

```
call_rainbow(  
  path,  
  format_in = c("agilent_d", "waters_raw", "masshunter", "chemstation", "chemstation_uv",  
    "chemstation_fid"),  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  by = c("detector", "name"),  
  what = NULL,  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw"),  
  collapse = TRUE,  
  precision = 1  
)
```

Arguments

path	Path to file
format_in	Format of the supplied files. Either agilent_d, waters_raw, or chemstation.
format_out	R format. Either matrix, data.frame, or data.table.

data_format	Whether to return data in wide or long format.
by	How to order the list that is returned. Either detector (default) or name.
what	What types of data to return (e.g. MS, UV, CAD, ELSD). This argument only applies if by == "detector".
read_metadata	Logical. Whether to attach metadata. Defaults to TRUE.
metadata_format	Format to output metadata. Either chromconverter or raw.
collapse	Logical. Whether to collapse lists that only contain a single element.
precision	Number of decimals to round mz values. Defaults to 1.

Value

Returns a (nested) list of matrices or data.frames according to the value of format_out. Data is ordered according to the value of by.

Author(s)

Ethan Bass

configure_openchrom *Configure OpenChrom parser*

Description

Configure OpenChrom parser

Usage

```
configure_openchrom(cli = c("null", "true", "false", "status"), path = NULL)
```

Arguments

cli	Defaults to NULL. If "true", R will rewrite openchrom ini file to enable CLI. If "false", R will disable CLI. If NULL, R will not modify the ini file.
path	Path to 'OpenChrom' executable (Optional). The supplied path will overwrite the current path.

Value

If cli is set to "status", returns a Boolean value indicating whether 'OpenChrom' is configured correctly. Otherwise, returns the path to OpenChrom command-line application.

Author(s)

Ethan Bass

extract_metadata	<i>Extract metadata</i>
------------------	-------------------------

Description

Extract metadata as a `data.frame` from a list of chromatograms.

Usage

```
extract_metadata(  
  chrom_list,  
  what = c("instrument", "detector", "detector_id", "software", "method", "batch",  
           "operator", "run_datetime", "sample_name", "sample_id", "injection_volume",  
           "time_range", "time_interval", "time_unit", "detector_range", "detector_y_unit",  
           "detector_x_unit", "intensity_multiplier", "scaled", "source_file",  
           "source_file_format", "source_sha1", "data_format", "parser", "format_out"),  
  format_out = c("data.frame", "tibble")  
)
```

Arguments

<code>chrom_list</code>	A list of chromatograms with attached metadata (as returned by <code>read_chroms</code> with <code>read_metadata = TRUE</code>).
<code>what</code>	A character vector specifying the metadata elements to extract.
<code>format_out</code>	Format of object. Either <code>data.frame</code> or <code>tibble</code> .

Value

A `data.frame` or `tibble` (according to the value of `format_out`), with samples as rows and the specified metadata elements as columns.

<code>read_agilent_d</code>	<i>Read files from 'ChemStation' .D directories</i>
-----------------------------	---

Description

Agilent `.ch` files come in several different varieties. This parser can automatically detect and read several versions of these files from 'Agilent ChemStation' and 'Agilent OpenLab', including versions 30 and 130, which are generally produced by ultraviolet detectors, as well as 81, 179, and 181 which are generally produced by flame ionization (FID) detectors.

Usage

```
read_agilent_d(
  path,
  what = c("chroms", "dad", "peak_table"),
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  collapse = TRUE
)
```

Arguments

path	Path to .ch file
what	Whether to extract chromatograms (chroms), DAD data (dad) and/or peak tables peak_table. Accepts multiple arguments. ms_spectra. Accepts multiple arguments.
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
collapse	Logical. Whether to collapse lists that only contain a single element.

Value

A list of chromatograms in the format specified by data_format and #' format_out. If data_format is wide, the chromatograms will be returned with retention times as rows and columns containing signal intensity for each signal. If long format is requested, retention times will be in the first column. The format_out argument determines whether the chromatogram is returned as a matrix, data.frame or data.table. Metadata can be attached to the chromatogram as [attributes](#) if read_metadata is TRUE.

Author(s)

Ethan Bass

read_agilent_dx

Parser for reading Agilent (.dx) files into R

Description

Parser for reading Agilent (.dx) files into R

Usage

```
read_agilent_dx(  
  path,  
  path_out = NULL,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE  
)
```

Arguments

path	Path to .dx file.
path_out	Path to directory to export unzipped files.
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
read_metadata	Logical. Whether to attach metadata.

Value

A chromatogram in the format specified by format_out (retention time x wavelength).

Author(s)

Ethan Bass

read_asm

Read Allotrope Simple Model (ASM) 2D chromatograms

Description

A parser for reading 'Allotrope Simple Model' files into R.

Usage

```
read_asm(  
  path,  
  data_format = c("wide", "long"),  
  format_out = c("matrix", "data.frame", "data.table"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw"),  
  collapse = TRUE  
)
```

Arguments

path	Path to ASM .json file.
data_format	Whether to return data in wide or long format.
format_out	Matrix or data.frame.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
collapse	Logical. Whether to collapse lists that only contain a single element. Defaults to TRUE.

Value

A 2D chromatogram in the format specified by `data_format` and `format_out`. If `data_format` is wide, the chromatogram will be returned with retention times as rows and a single column for the intensity. If long format is requested, two columns will be returned: one for the retention time and one for the intensity. The `format_out` argument determines whether the chromatogram is returned as a matrix or data.frame. Metadata can be attached to the chromatogram as `attributes` if `read_metadata` is TRUE.

Author(s)

Ethan Bass

read_cdf	<i>Read CDF file</i>
----------	----------------------

Description

Parser for Analytical Data Interchange (ANDI) netCDF files.

Usage

```
read_cdf(
  path,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  what = NULL,
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  collapse = TRUE,
  ...
)
```

Arguments

path	Path to ANDI netCDF file.
format_out	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .
data_format	Whether to return data in wide or long format. For 2D files, "long" format returns the retention time as the first column of the <code>data.frame</code> or <code>matrix</code> while "wide" format returns the retention time as the rownames of the object.
what	For ANDI chrom files, whether to extract choms and/or <code>peak_table</code> . For ANDI ms files, whether to extract MS1 scans (MS1) or the total ion chromatogram (TIC).
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .
collapse	Logical. Whether to collapse lists that only contain a single element.
...	Additional arguments to parser. The <code>ms_format</code> argument can be used here to specify whether to return mass spectra in <code>list</code> format or as a <code>data.frame</code> .

Value

A chromatogram in the format specified by the `format_out` and `data_format` arguments.

Author(s)

Ethan Bass

read_chemstation_ch *Read 'ChemStation' CH files*

Description

Agilent `.ch` files come in several different varieties. This parser can automatically detect and read several versions of these files from 'Agilent ChemStation' and 'Agilent OpenLab', including versions 30 and 130, which are generally produced by ultraviolet detectors, as well as 81, 179, and 181 which are generally produced by flame ionization (FID) detectors.

Usage

```
read_chemstation_ch(
  path,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  scale = TRUE
)
```

Arguments

path	Path to .ch file
format_out	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .
data_format	Whether to return data in wide or long format.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .
scale	Whether to scale the data by the scaling factor present in the file. Defaults to <code>TRUE</code> . 'MassHunter' seems to ignore the scaling factor in at least some types of 'ChemStation' files.

Value

A 2D chromatogram in the format specified by `data_format` and `format_out`. If `data_format` is wide, the chromatogram will be returned with retention times as rows and a single column for the intensity. If long format is requested, two columns will be returned: one for the retention time and one for the intensity. The `format_out` argument determines whether the chromatogram is returned as a `matrix`, `data.frame`, or `data.table`. Metadata can be attached to the chromatogram as `attributes` if `read_metadata` is `TRUE`.

Note

This function was adapted from the [Chromatography Toolbox](#) (© James Dillon 2014).

Author(s)

Ethan Bass

read_chemstation_csv *Read Chemstation CSV*

Description

Reads 'Agilent Chemstation' CSV files into R.

Usage

```
read_chemstation_csv(
  path,
  format_out = c("matrix", "data.frame", "data.table")
)
```

Arguments

path	Path to file
format_out	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .

Details

'Agilent Chemstation' CSV files are encoded in UTF-16.

Value

A chromatogram in the format specified by `format_out` (retention time x wavelength).

Author(s)

Ethan Bass

`read_chemstation_ms` *Read 'Agilent ChemStation' MS file.*

Description

Parser to read Agilent Chemstation 'MSD Spectral Files' beginning with `x01/x32/x00/x00`.

Usage

```
read_chemstation_ms(
  path,
  what = c("MS1", "BPC", "TIC"),
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  collapse = TRUE
)
```

Arguments

<code>path</code>	Path to .ms file
<code>what</code>	What stream to get: current options are MS1, BPC and/or TIC. If a stream is not specified, the function will return all streams.
<code>format_out</code>	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .
<code>data_format</code>	Whether to return data in wide or long format.
<code>read_metadata</code>	Logical. Whether to attach metadata.
<code>metadata_format</code>	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .
<code>collapse</code>	Logical. Whether to collapse lists that only contain a single element. Defaults to TRUE.

Value

A 2D chromatogram in the format specified by `data_format` and `format_out`. If `data_format` is wide, the chromatogram will be returned with retention times as rows and a single column for the intensity. If long format is requested, two columns will be returned: one for the retention time and one for the intensity. The `format_out` argument determines whether the chromatogram is returned as a `matrix`, `data.frame`, or `data.table`. Metadata can be attached to the chromatogram as `attributes` if `read_metadata` is `TRUE`.

Note

Many thanks to Evan Shi and Eugene Kwan for providing helpful information on the structure of these files in the [rainbow documentation](#).

Author(s)

Ethan Bass

read_chemstation_reports

Read 'Agilent ChemStation' Reports

Description

Read 'Agilent ChemStation' Reports

Usage

```
read_chemstation_reports(  
  paths,  
  data_format = c("chromatographr", "original"),  
  metadata_format = c("chromconverter", "raw")  
)
```

Arguments

<code>paths</code>	Paths to 'ChemStation' report files.
<code>data_format</code>	Format to output data. Either <code>chromatographr</code> or <code>chemstation</code> .
<code>metadata_format</code>	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .

Value

A `data.frame` containing the information from the specified 'ChemStation' report.

Author(s)

Ethan Bass

read_chemstation_uv *Read 'ChemStation' DAD files*

Description

Agilent .uv files come in several different varieties. This parser can automatically detect and read several versions of these files from 'Agilent ChemStation' and 'Agilent OpenLab', including versions 31 and 131.

Usage

```
read_chemstation_uv(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw"),  
  scale = TRUE  
)
```

Arguments

path	Path to .uv file.
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Either wide (default) or long.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
scale	Whether to scale the data by the scaling factor present in the file. Defaults to TRUE.

Value

A 3D chromatogram in the format specified by data_format and format_out. If data_format is wide, the chromatogram will be returned with retention times as rows and wavelengths as columns. If long format is requested, three columns will be returned: one for the retention time, one for the wavelength and one for the intensity. The format_out argument determines whether the chromatogram is returned as a matrix or data.frame. Metadata can be attached to the chromatogram as [attributes](#) if read_metadata is TRUE.

Note

This function was adapted from the parser in the rainbow project licensed under GPL 3 by Evan Shi <https://rainbow-api.readthedocs.io/en/latest/agilent/uv.html>.

Author(s)

Ethan Bass

read_chromeleon	<i>Chromeleon ASCII reader</i>
-----------------	--------------------------------

Description

Reads 'Thermo Fisher Chromeleon™ CDS' files into R.

Usage

```
read_chromeleon(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw")  
)
```

Arguments

path	Path to file
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either chromconverter or raw.

Value

A chromatogram in the format specified by format_out. (retention time x wavelength).

Author(s)

Ethan Bass

read_chroms	<i>Read Chromatograms</i>
-------------	---------------------------

Description

Reads chromatograms from specified folders or vector of paths using either an internal parser or bindings to an external library, such as [Aston](#), [Entab](#), [ThermoRawFileParser](#), [OpenChrom](#), [rainbow](#).

Usage

```
read_chroms(
  paths,
  format_in = c("agilent_d", "agilent_dx", "asm", "chemstation", "chemstation_fid",
    "chemstation_ch", "chemstation_csv", "chemstation_ms", "chemstation_uv",
    "masshunter_dad", "chromeleon_uv", "mzml", "mzxml", "mdf", "shimadzu_ascii",
    "shimadzu_dad", "shimadzu_fid", "shimadzu_gcd", "shimadzu_qgd", "shimadzu_lcd",
    "thermoraw", "varian_sms", "waters_arw", "waters_raw", "msd", "csd", "wsd", "other"),
  find_files,
  pattern = NULL,
  parser = c("", "chromconverter", "aston", "entab", "thermoraw", "openchrom", "rainbow"),
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  path_out = NULL,
  export_format = c("", "csv", "chemstation_csv", "cdf", "mzml", "animl"),
  force = FALSE,
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  progress_bar,
  cl = 1,
  verbose = getOption("verbose"),
  sample_names = NULL,
  dat = NULL,
  ...
)
```

Arguments

paths	Paths to data files or directories containing the files.
format_in	Format of files to be imported/converted. Current options include: agilent_d, agilent_dx, chemstation, chemstation_uv, chemstation_ch, chemstation_csv, chemstation_ms, masshunter, masshunter_dad, chromeleon_uv, shimadzu_ascii, shimadzu_fid, shimadzu_dad, thermoraw, waters_arw, waters_raw, mzml, mzxml, cdf, mdf, msd, csd, wsd, or other.
find_files	Logical. Set to TRUE (default) if you are providing the function with a folder or vector of folders containing the files. Otherwise, set to FALSE.

pattern	pattern (e.g. a file extension). Defaults to NULL, in which case file extension will be deduced from format_in.
parser	What parser to use (optional). Current options are chromconverter, aston, entab, thermoraw, openchrom, or rainbow.
format_out	Class of output. Either matrix, data.frame, or data.table .
data_format	Whether to output data in wide or long format. Either wide or long.
path_out	Path for exporting files. If path not specified, files will export to current working directory.
export_format	Export format. Currently the options include .csv, chemstation_csv (utf-16 encoding), and cdf, unless you are using OpenChrom parsers, where there are two additional options: mzml, and animl.
force	Logical. Whether to overwrite files when exporting. Defaults to FALSE.
read_metadata	Logical, whether to attach metadata (if it's available). Defaults to TRUE.
metadata_format	Format to output metadata. Either chromconverter or raw.
progress_bar	Logical. Whether to show progress bar. Defaults to TRUE if pbapply is installed.
cl	Argument to pbapply specifying the number of clusters to use or a cluster object created by makeCluster . Defaults to 1.
verbose	Logical. Whether to print output from external parsers to the R console.
sample_names	An optional character vector of sample names. Otherwise sample names default to the basename of the specified files.
dat	Existing list of chromatograms to append results. (Defaults to NULL).
...	Additional arguments to parser.

Details

Provides a unified interface to all chromConverter parsers. Currently recognizes 'Agilent ChemStation' (.uv, .ch, .dx), 'Agilent MassHunter' (.dad), 'Thermo RAW' (.raw), 'Waters ARW' (.arw), 'Waters RAW' (.raw), 'Chromeleon ASCII' (.txt), 'Shimadzu ASCII' (.txt), 'Shimadzu GCD', 'Shimadzu LCD' (DAD and chromatogram streams) and 'Shimadzu QGD' files. Also, wraps 'OpenChrom' parsers, which include many additional formats. To use 'Entab', 'ThermoRaw-FileParser', or 'OpenChrom' parsers, they must be manually installed. Please see the instructions in the [README](#) for further details.

If paths to individual files are provided, read_chroms will try to infer the file format and select an appropriate parser. However, when providing paths to directories, the file format must be specified using the format_in argument.

Value

A list of chromatograms in matrix, data.frame, or data.table format, according to the value of format_out. Chromatograms may be returned in either wide or long format according to the value of data_format.

Side effects

If an `export_format` is provided, chromatograms will be exported in the specified format specified into the folder specified by `path_out`. Files can currently be converted to `csv`, `mzml`, or `cdf` format. If an `openchrom` parser is selected, `ANIML` is available as an additional option.

Author(s)

Ethan Bass

Examples

```
path <- "tests/testthat/testdata/dad1.uv"
chr <- read_chroms(path, find_files = FALSE, format_in = "chemstation_uv")
```

read_mdf

Read 'Lumex' MDF files into R

Description

Read 'Lumex' MDF files into R

Usage

```
read_mdf(
  path,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE
)
```

Arguments

<code>path</code>	The path to a 'Lumex' .mdf file.
<code>format_out</code>	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .
<code>data_format</code>	Whether to return data in wide or long format.
<code>read_metadata</code>	Whether to read metadata from file.

Value

A chromatogram in the format specified by the `format_out` and `data_format` arguments (retention time x wavelength).

Author(s)

Ethan Bass

read_mzml	<i>Extract data from mzML files</i>
-----------	-------------------------------------

Description

Extracts data from mzML files using parsers from either RaMS or mzR. The RaMS parser (default) will only return data in tidy (long) format. The mzR parser will return data in wide format. Currently the mzR-based parser is configured to return only DAD data.

Usage

```
read_mzml(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("long", "wide"),  
  parser = c("RaMS", "mzR"),  
  what = c("MS1", "MS2", "BPC", "TIC", "DAD", "chroms", "metadata", "everything"),  
  verbose = FALSE,  
  ...  
)
```

Arguments

path	path to file
format_out	Class of output. Only applies if mzR is selected. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> . RaMS will return a list of <code>data.tables</code> regardless of what is selected here.
data_format	Whether to return data in wide or long format.
parser	What parser to use. Either RaMS or mzR.
what	What types of data to return (argument to grabMSdata . Options include MS1, MS2, BPC, TIC, DAD, chroms, metadata, or everything).
verbose	Argument to grabMSdata controlling verbosity.
...	Additional arguments to grabMSdata .

Value

If RaMS is selected, the function will return a list of "tidy" `data.table` objects. If mzR is selected, the function will return a chromatogram in `matrix` or `data.frame` format according to the value of `format_out`.

Author(s)

Ethan Bass

read_peaklist	<i>Read peak lists</i>
---------------	------------------------

Description

Reads peak lists from specified folders or vector of paths.

Usage

```
read_peaklist(  
  paths,  
  find_files,  
  format_in = c("chemstation", "shimadzu_fid", "shimadzu_dad"),  
  pattern = NULL,  
  data_format = c("chromatographr", "original"),  
  metadata_format = c("chromconverter", "raw"),  
  read_metadata = TRUE,  
  progress_bar,  
  cl = 1  
)
```

Arguments

paths	paths to files or folders containing files.
find_files	Logical. Set to TRUE (default) if you are providing the function with a folder or vector of folders containing the files. Otherwise, set to FALSE.
format_in	Format of files to be imported/converted. Current options include: chemstation or shimadzu.
pattern	pattern (e.g. a file extension). Defaults to NULL, in which case file extension will be deduced from format_in.
data_format	Either chromatographr or original.
metadata_format	Format to output metadata. Either chromconverter or raw.
read_metadata	Logical, whether to attach metadata (if it's available). Defaults to TRUE.
progress_bar	Logical. Whether to show progress bar. Defaults to TRUE if pbapply is installed.
cl	Argument to pbapply specifying the number of clusters to use or a cluster object created by makeCluster . Defaults to 1.

Value

A list of chromatograms in matrix or data.frame format, according to the value of format_out.

Author(s)

Ethan Bass

Examples

```
path <- "tests/testthat/testdata/dad1.uv"
chr <- read_chroms(path, find_files = FALSE, format_in = "chemstation_uv")
```

read_shimadzu	<i>Shimadzu ascii reader</i>
---------------	------------------------------

Description

Reads 'Shimadzu' ascii files into R. These files can be exported from 'Shimadzu LabSolutions' by right clicking on samples in the sample list and selecting File Conversion:Convert to ASCII.

Usage

```
read_shimadzu(
  path,
  what = "chroms",
  format_in = NULL,
  include = c("fid", "lc", "dad", "uv", "tic"),
  format_out = c("matrix", "data.frame"),
  data_format = c("wide", "long"),
  peaktable_format = c("chromatographr", "original"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  ms_format = c("data.frame", "list"),
  collapse = TRUE,
  scale = TRUE
)
```

Arguments

path	Path to file.
what	Whether to extract chromatograms (chroms), peak_table, and/or ms_spectra. Accepts multiple arguments.
format_in	This argument is deprecated and is no longer required.
include	Which chromatograms to include. Options are fid, dad, uv, tic, and status.
format_out	R format. Either matrix or data.frame.
data_format	Whether to return data in wide or long format.
peaktable_format	Whether to return peak tables in chromatographr or original format.
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either chromconverter or raw.

ms_format	Whether to return mass spectral data as a (long) <code>data.frame</code> or a list.
collapse	Logical. Whether to collapse lists that only contain a single element.
scale	Whether to scale the data by the scaling factor present in the file. Defaults to TRUE.

Value

A nested list of elements from the specified file, where the top levels are chromatograms, peak tables, and/or mass spectra according to the value of `what`. Chromatograms are returned in the format specified by `format_out` (retention time x wavelength).

Author(s)

Ethan Bass

read_shimadzu_gcd *Shimadzu GCD parser*

Description

Read chromatogram data streams from 'Shimadzu' GCD files.

Usage

```
read_shimadzu_gcd(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw")  
)
```

Arguments

path	Path to GCD file.
format_out	Class of output. Either <code>matrix</code> , <code>data.frame</code> , or <code>data.table</code> .
data_format	Either <code>wide</code> (default) or <code>long</code> .
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .

Details

A parser to read chromatogram data streams from 'Shimadzu' .gcd files. GCD files are encoded as 'Microsoft' OLE documents. The parser relies on the `olefile` package in Python to unpack the files. The PDA data is encoded in a stream called PDA 3D Raw Data: 3D Raw Data. The GCD data stream contains a segment for each retention time, beginning with a 24-byte header.

The 24 byte header consists of the following fields:

- 4 bytes: segment label (17234).
- 4 bytes: Little-endian integer specifying the sampling interval in milliseconds.
- 4 bytes: Little-endian integer specifying the number of values in the file.
- 4 bytes: Little-endian integer specifying the total number of bytes in the file (However, this seems to be off by a few bytes?).
- 8 bytes of `00s`

After the header, the data are simply encoded as 64-bit (little-endian) floating-point numbers. The retention times can be (approximately?) derived from the number of values and the sampling interval encoded in the header.

Value

A 2D chromatogram from the chromatogram stream in `matrix` or `data.frame` format, according to the value of `format_out`. The chromatograms will be returned in `wide` or `long` format according to the value of `data_format`.

Note

This parser is experimental and may still need some work. It is not yet able to interpret much metadata from the files.

Author(s)

Ethan Bass

read_shimadzu_lcd *Shimadzu LCD parser*

Description

Read 3D PDA or 2D chromatogram streams from 'Shimadzu' LCD files.

Usage

```
read_shimadzu_lcd(
  path,
  what,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw"),
  scale = TRUE,
  collapse = TRUE
)
```

Arguments

path	Path to LCD file.
what	What stream to get: current options are pda, chromatograms chrams, or tic. If a stream is not specified, the function will default to pda if the PDA stream is present.
format_out	Matrix or data.frame.
data_format	Either wide (default) or long.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
scale	Whether to scale the data by the value factor. Defaults to TRUE.
collapse	Logical. Whether to collapse lists that only contain a single element.

Details

A parser to read data from 'Shimadzu' .lcd files. LCD files are encoded as 'Microsoft' OLE documents. The parser relies on the `olefile` package in Python to unpack the files. The PDA data is encoded in a stream called PDA 3D Raw Data: 3D Raw Data. The PDA data stream contains a segment for each retention time, beginning with a 24-byte header.

The 24 byte header consists of the following fields:

- 4 bytes: segment label (17234).
- 4 bytes: Little-endian integer specifying the sampling rate along the time axis for 2D streams or along the spectral axis (?) for PDA streams.
- 4 bytes: Little-endian integer specifying the number of values in the file (for 2D data) or the number of wavelength values in each segment (for 3D data).
- 4 bytes: Little-endian integer specifying the total number of bytes in the segment.
- 8 bytes of 00.

For 3D data, Each time point is divided into two sub-segments, which begin and end with an integer specifying the length of the sub-segment in bytes. 2D data are structured similarly but with more segments. All known values in this the LCD data streams are little-endian and the data are delta-encoded. The first hexadecimal digit of each value is a sign digit specifying the number of bytes

in the delta and whether the value is positive or negative. The sign digit represents the number of hexadecimal digits used to encode each value. Even numbered sign digits correspond to positive deltas, whereas odd numbers indicate negative deltas. Positive values are encoded as little-endian integers, while negative values are encoded as two's complements. The value at each position is derived by subtracting the delta at each position from the previous value.

Value

A chromatogram or list of chromatograms in the format specified by `data_format` and `format_out`. If `data_format` is wide, the chromatogram(s) will be returned with retention times as rows and a single column for the intensity. If long format is requested, two columns will be returned: one for the retention time and one for the intensity. The `format_out` argument determines whether chromatograms are returned as a matrix, `data.frame`, or `data.table`. Metadata can be attached to the chromatogram as `attributes` if `read_metadata` is TRUE.

Note

My parsing of the date-time format seems to be a little off, since the acquisition times diverge slightly from the ASCII file.

Author(s)

Ethan Bass

read_shimadzu_qgd *Read 'Shimadzu' QGD*

Description

Parser for 'Shimadzu GCMSsolution' .qgd files.

Usage

```
read_shimadzu_qgd(  
  path,  
  what = c("MS1", "TIC"),  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw"),  
  collapse = TRUE  
)
```

Arguments

path	Path to 'Shimadzu' QGD file.
what	What stream to get: current options are MS1 and/or TIC. If a stream is not specified, the function will return both streams.
format_out	Matrix or data.frame.
data_format	Either wide (default) or long.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
collapse	Logical. Whether to collapse lists that only contain a single element. Defaults to TRUE.

Details

The MS data is stored in the "GCMS Raw Data" storage, which contains a MS Raw Data stream with MS scans, a TIC Data stream containing the total ion chromatogram, and a Retention Time stream containing the retention times. All known values are little-endian. The retention time stream is a simple array of 4-byte integers. The TIC stream is a simple array of 8-byte integers corresponding to retention times stored in the retention time stream. The MS Raw Data stream is blocked by retention time. Each block begins with a header consisting of the following elements:

- scan number (4-byte integer)
- retention time (4-byte integer)
- unknown (12-bytes)
- number of bytes in intensity values (2-byte integer)
- unknown (8-bytes)

After the header, the rest of the block consists of an array of m/z values and intensities. The m/z values are encoded as 2-byte integers where each m/z value is scaled by a factor of 20. Intensities are encoded as (unsigned) integers with variable byte-length defined by the value in the header.

Value

A 2D chromatogram from the chromatogram stream in `matrix`, `data.frame`, or `data.table` format, according to the value of `format_out`. The chromatograms will be returned in wide or long format according to the value of `data_format`.

A chromatogram or list of chromatograms in the format specified by `data_format` and `format_out`. If `data_format` is wide, the chromatogram(s) will be returned with retention times as rows and a single column for the intensity. If long format is requested, two columns will be returned: one for the retention time and one for the intensity. The `format_out` argument determines whether chromatograms are returned as a `matrix`, `data.frame`, or `data.table`. Metadata can be attached to the chromatogram as `attributes` if `read_metadata` is TRUE.

Note

This parser is experimental and may still need some work. It is not yet able to interpret much metadata from the files.

Author(s)

Ethan Bass

read_sz_lcd_2d	<i>Shimadzu LCD 2D parser</i>
----------------	-------------------------------

Description

Read 2D PDA data stream from 'Shimadzu' LCD files.

Usage

```
read_sz_lcd_2d(  
    path,  
    format_out = "data.frame",  
    data_format = "wide",  
    read_metadata = TRUE,  
    metadata_format = "shimadzu_lcd",  
    scale = TRUE  
)
```

Arguments

path	Path to LCD file.
format_out	Matrix or data.frame.
data_format	Either wide (default) or long.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.
scale	Whether to scale the data by the value factor.

Details

A parser to read chromatogram data streams from 'Shimadzu' .lcd files. LCD files are encoded as 'Microsoft' OLE documents. The parser relies on the [olefile](#) package in Python to unpack the files. The chromatogram data is encoded in streams titled LSS Raw Data:Chromatogram Ch<#>. The chromatogram data streams begin with a 24-byte header.

The 24 byte header consists of the following fields:

- 4 bytes: segment label (17234).
- 4 bytes: Little-endian integer specifying the sampling rate (in milliseconds).
- 4 bytes: Little-endian integer specifying the number of values in the file.
- 4 bytes: Little-endian integer specifying the total number of bytes in the file.
- 8 bytes of 00s

Each segment is divided into multiple sub-segments, which begin and end with an integer specifying the length of the sub-segment in bytes. All known values in this data stream are little-endian and the data are delta-encoded. The first hexadecimal digit of each value is a sign digit specifying the number of bytes in the delta and whether the value is positive or negative. The sign digit represents the number of hexadecimal digits used to encode each value. Even numbered sign digits correspond to positive deltas, whereas odd numbers indicate negative deltas. Positive values are encoded as little-endian integers, while negative values are encoded as two's complements. The value at each position is derived by subtracting the delta at each position from the previous value.

Value

One or more 2D chromatograms from the chromatogram streams in `matrix` or `data.frame` format, according to the value of `format_out`. If multiple chromatograms are found, they will be returned as a list of matrices or `data.frames`. The chromatograms will be returned in wide or long format according to the value of `data_format`.

Author(s)

Ethan Bass

read_sz_lcd_3d	<i>Shimadzu LCD 3D parser</i>
----------------	-------------------------------

Description

Read 3D PDA data stream from 'Shimadzu' LCD files.

Usage

```
read_sz_lcd_3d(
  path,
  format_out = "matrix",
  data_format = "wide",
  read_metadata = TRUE,
  metadata_format = "shimadzu_lcd",
  scale = TRUE
)
```

Arguments

<code>path</code>	Path to LCD file.
<code>format_out</code>	Matrix or <code>data.frame</code> .
<code>data_format</code>	Either wide (default) or long.
<code>read_metadata</code>	Logical. Whether to attach metadata.
<code>metadata_format</code>	Format to output metadata. Either <code>chromconverter</code> or <code>raw</code> .
<code>scale</code>	Whether to scale the data by the value factor.

Details

A parser to read PDA data from 'Shimadzu' .lcd files. LCD files are encoded as 'Microsoft' OLE documents. The parser relies on the [olefile](#) package in Python to unpack the files. The PDA data is encoded in a stream called PDA 3D Raw Data: 3D Raw Data. The PDA data stream contains a segment for each retention time, beginning with a 24-byte header.

The 24 byte header consists of the following fields:

- 4 bytes: segment label (17234).
- 4 bytes: Little-endian integer specifying the wavelength bandwidth (?).
- 4 bytes: Little-endian integer specifying the number of wavelength values in the segment.
- 4 bytes: Little-endian integer specifying the total number of bytes in the segment.
- 8 bytes of 00s

Each segment is divided into two sub-segments, which begin and end with an integer specifying the length of the sub-segment in bytes. All known values in this data stream are little-endian and the data are delta-encoded. The first hexadecimal digit of each value is a sign digit specifying the number of bytes in the delta and whether the value is positive or negative. The sign digit represents the number of hexadecimal digits used to encode each value. Even numbered sign digits correspond to positive deltas, whereas odd numbers indicate negative deltas. Positive values are encoded as little-endian integers, while negative values are encoded as two's complements. The value at each position is derived by subtracting the delta at each position from the previous value.

Value

A 3D chromatogram from the PDA stream in `matrix` or `data.frame` format, according to the value of `format_out`. The chromatograms will be returned in `wide` or `long` format according to the value of `data_format`.

Author(s)

Ethan Bass

read_thermoraw

Read ThermoRaw files into R using ThermoRawFileParser

Description

Converts ThermoRawFiles to mzML by calling the [ThermoRawFileParser](#) from the command-line.

Usage

```
read_thermoraw(  
  path,  
  path_out = NULL,  
  format_out = c("matrix", "data.frame"),  
  read_metadata = TRUE,
```

```
    metadata_format = c("chromconverter", "raw"),
    verbose = getOption("verbose")
  )
```

Arguments

path	Path to file.
path_out	Path to directory to export mzML files. If path_out isn't specified, a temp directory will be used.
format_out	R format. Either matrix or data.frame.
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either chromconverter or raw.
verbose	Logical. Whether to print output from OpenChrom to the console.

Details

To use this function, the ThermoRawFileParser must be manually installed.

Value

A chromatogram in the format specified by format_out.

Side effects

Exports chromatograms in mzML format to the folder specified by path_out.

Author(s)

Ethan Bass

References

Hulstaert Niels, Jim Shofstahl, Timo Sachsenberg, Mathias Walzer, Harald Barsnes, Lennart Martens, and Yasset Perez-Riverol. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *Journal of Proteome Research* **19**, no. 1 (January 3, 2020): 537–42. doi:[10.1021/acs.jproteome.9b00328](https://doi.org/10.1021/acs.jproteome.9b00328).

Examples

```
## Not run:
read_thermoraw(path)

## End(Not run)
```

read_varian_peaklist *Read 'Varian' peaklist. Read peak list(s) from 'Varian MS Workstation'.*

Description

Read 'Varian' peaklist. Read peak list(s) from 'Varian MS Workstation'.

Usage

```
read_varian_peaklist(path)
```

Arguments

path Path to Varian peak list file.

Value

A data.frame containing the information from the specified report.

Author(s)

Ethan Bass

read_varian_sms *Read 'Varian' SMS*

Description

Parser for reading 'Varian Workstation' SMS file.

Usage

```
read_varian_sms(  
  path,  
  what = c("MS1", "TIC", "BPC"),  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  collapse = TRUE  
)
```


Arguments

path	Path to .SMS files.
what	Whether to extract chromatograms (chroms) and/or MS1 data. Accepts multiple arguments.
format_out	R format. Either <code>matrix</code> or <code>data.frame</code> .
data_format	Whether to return data in wide or long format.
read_metadata	Whether to read metadata from file. (This is just a placeholder for now as there is not yet support for parsing metadata).
collapse	Logical. Whether to collapse lists that only contain a single element.

Details

Varian SMS files begin with a "DIRECTORY" with offsets for each section. The first section (in all the files I've been able to inspect) is "MSData" generally beginning at byte 3238. This MSData section is in turn divided into two sections. The first section (after a short header) contains chromatogram data. Some of the information found in this section includes scan numbers, retention times, (as 64-bit floats), the total ion chromatogram (TIC), the base peak chromatogram (BPC), ion time (μ sec), as well as some other unidentified information. The scan numbers and intensities for the TIC and BPC are stored at 4-byte little-endian integers. Following this section, there is a series of null bytes, followed by a series of segments containing the mass spectra.

The encoding scheme for the mass spectra is somewhat more complicated. Each scan is represented by a series of values of variable length separated from the next scan by two null bytes. Within these segments, values are paired. The first value in each pair represents the delta-encoded mass-to-charge ratio, while the second value represents the intensity of the signal. Values in this section are variable-length, big-endian integers that are encoded using a selective bit masking based on the leading digit (d) of each value. The length of each integer seems to be determined as $1 + (d \% 4)$. Integers beginning with digits 0-3 are simple 2-byte integers. If $d \geq 4$, values are determined by masking to preserve the lowest n bits according to the following scheme:

- d = 4-5 -> preserve lowest 13 bits
- d = 6-7 -> preserve lowest 14 bits
- d = 8-9 -> preserve lowest 21 bits
- d = 10-11 (A-B) -> preserve lowest 22 bits
- d = 12-13 (C-D) -> preserve lowest 27 bits
- d = 14-15 (E-F) -> preserve lowest 28 bits (?)

Value

A chromatogram or list of chromatograms from the specified file, according to the value of `what`. Chromatograms are returned in the format specified by `format_out`.

Note

There is still only limited support for the extraction of metadata from this file format. Also, the timestamp conversions aren't quite right.

Author(s)

Ethan Bass

read_waters_arw	<i>Waters ascii (.arw) reader</i>
-----------------	-----------------------------------

Description

Reads 'Waters ARW' files.

Usage

```
read_waters_arw(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw")  
)
```

Arguments

path	Path to file
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
read_metadata	Whether to read metadata from file.
metadata_format	Format to output metadata. Either chromconverter or raw.

DetailsFor help exporting files from Empower, you can consult the official documentation: [How_to_export_3D_raw_data_from_Empower](#)**Value**

A chromatogram in the format specified by format_out (retention time x wavelength).

Author(s)

Ethan Bass

read_waters_raw	<i>Read 'Waters' RAW</i>
-----------------	--------------------------

Description

Parser for reading 'Waters MassLynx (.raw) files into R.

Usage

```
read_waters_raw(  
  path,  
  format_out = c("matrix", "data.frame", "data.table"),  
  data_format = c("wide", "long"),  
  read_metadata = TRUE,  
  metadata_format = c("chromconverter", "raw")  
)
```

Arguments

path	Path to .raw file.
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Either wide (default) or long.
read_metadata	Logical. Whether to attach metadata.
metadata_format	Format to output metadata. Either chromconverter or raw.

Value

A chromatogram in the format specified by format_out (retention time x wavelength).

Note

For now this parser only reads 1D chromatograms (not mass spectra or DAD data) and does not support parsing of metadata from 'Waters' RAW files.

Author(s)

Ethan Bass

sp_converter	<i>Converter for 'Agilent MassHunter' UV files</i>
--------------	--

Description

Converts a single chromatogram from MassHunter .sp format to R data.frame.

Usage

```
sp_converter(
  path,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw")
)
```

Arguments

path	Path to file
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
read_metadata	Logical. Whether to read metadata and attach it to the chromatogram.
metadata_format	Format to output metadata. Either chromconverter or raw.

Details

Uses the [Aston](#) file parser.

Value

A chromatogram in data.frame format (retention time x wavelength).

uv_converter	<i>Converter for 'Agilent ChemStation' UV files</i>
--------------	---

Description

Converts a single chromatogram from ChemStation .uv format to R data.frame.

Usage

```
uv_converter(
  path,
  format_out = c("matrix", "data.frame", "data.table"),
  data_format = c("wide", "long"),
  correction = TRUE,
  read_metadata = TRUE,
  metadata_format = c("chromconverter", "raw")
)
```

Arguments

path	Path to file
format_out	Class of output. Either matrix, data.frame, or data.table.
data_format	Whether to return data in wide or long format.
correction	Logical. Whether to apply empirical correction. Defaults is TRUE.
read_metadata	Logical. Whether to read metadata and attach it to the chromatogram.
metadata_format	Format to output metadata. Either chromconverter or raw.

Details

Uses the [Aston](#) file parser.

Value

A chromatogram in data.frame format (retention time x wavelength).

write_andi_chrom	<i>Write ANDI chrom CDF file from chromatogram</i>
------------------	--

Description

Exports a chromatogram in ANDI (Analytical Data Interchange) chromatography format (ASTM E1947-98). This format can only accommodate unidimensional data. For two-dimensional chromatograms, the column to export can be specified using the lambda argument. Otherwise, a warning will be generated and the first column of the chromatogram will be exported.

Usage

```
write_andi_chrom(x, path_out, sample_name = NULL, lambda = NULL, force = FALSE)
```

Arguments

x	A chromatogram in (wide) format.
path_out	The path to write the file.
sample_name	The name of the file. If a name is not provided, the name will be derived from the sample_name attribute.
lambda	The wavelength to export (for 2-dimensional chromatograms). Must be a string matching one the columns in x or the index of the column to export.
force	Whether to overwrite existing files at the specified path. Defaults to FALSE.

Value

No return value. The function is called for its side effects.

Side effects

Exports a chromatogram in ANDI chromatography format (netCDF) in the directory specified by path_out. The file will be named according to the value of sample_name. If no sample_name is provided, the sample_name attribute will be used if it exists.

Author(s)

Ethan Bass

write_chroms

Export chromatograms

Description

Export chromatograms

Usage

```
write_chroms(  
  chrom_list,  
  path_out,  
  export_format = c("mzml", "cdf", "csv"),  
  what = "",  
  force = FALSE,  
  show_progress = TRUE,  
  verbose = getOption("verbose"),  
  ...  
)
```

Arguments

chrom_list	A list of chromatograms.
path_out	Path to directory for writing files.
export_format	Format to export files: either mzml, cdf, or csv.
what	What to write. Either MS1 or chrom.
force	Logical. Whether to overwrite existing files. Defaults to TRUE.
show_progress	Logical. Whether to show progress bar. Defaults to TRUE.
verbose	Logical. Whether to print verbose output.
...	Additional arguments to write function.

Author(s)

Ethan Bass

write_mzml

Write mzML

Description

This function constructs mzML files using a streaming XML-writing approach, generating and writing XML content directly to a file connection. This method is memory-efficient and fast, making it well-suited for large mzML files. However, it may be less flexible than DOM-based methods for handling complex XML structures or large-scale modifications after writing has begun.

Usage

```
write_mzml(  
  data,  
  path_out,  
  sample_name = NULL,  
  what = NULL,  
  instrument_info = NULL,  
  compress = TRUE,  
  indexed = TRUE,  
  force = FALSE,  
  show_progress = TRUE,  
  verbose = getOption("verbose")  
)
```

Arguments

<code>data</code>	List of <code>data.frames</code> or <code>data.tables</code> containing spectral data.
<code>path_out</code>	The path to write the file.
<code>sample_name</code>	The name of the file. If a name is not provided, the name will be derived from the <code>sample_name</code> attribute.
<code>what</code>	Which streams to write to mzML: "ms1", "ms2", "tic", "bpc", and/or "dad".
<code>instrument_info</code>	Instrument info to write to mzML file.
<code>compress</code>	Logical. Whether to use zlib compression. Defaults to TRUE.
<code>indexed</code>	Logical. Whether to write indexed mzML. Defaults to TRUE.
<code>force</code>	Logical. Whether to overwrite existing files at <code>path_out</code> . Defaults to FALSE.
<code>show_progress</code>	Logical. Whether to show progress bar. Defaults to TRUE.
<code>verbose</code>	Logical. Whether or not to print status messages.

Details

This function constructs mzML files by writing XML strings directly to a file connection. While this approach is memory-efficient and fast, it may be less flexible for handling complex XML structures.

The function supports writing various types of spectral data including MS1, MS2, TIC (Total Ion Current), BPC (Base Peak Chromatogram), and DAD (Diode Array Detector) data.

If `indexed = TRUE`, the function will create an indexed mzML file, which allows faster random access to spectra.

Value

Invisibly returns the path to the written mzML file.

Author(s)

Ethan Bass

Index

attributes, [8](#), [10](#), [12](#), [14](#), [15](#), [26](#), [27](#)

basename, [18](#)

call_entab, [3](#)
call_openchrom, [3](#)
call_rainbow, [5](#)
configure_openchrom, [6](#)

data.table, [11](#), [18](#)

extract_metadata, [7](#)

grabMSdata, [20](#)

makeCluster, [18](#), [21](#)

pbapply, [18](#), [21](#)

read_agilent_d, [7](#)
read_agilent_dx, [8](#)
read_asm, [9](#)
read_cdf, [10](#)
read_chemstation_ch, [11](#)
read_chemstation_csv, [12](#)
read_chemstation_ms, [13](#)
read_chemstation_reports, [14](#)
read_chemstation_uv, [15](#)
read_chromeleon, [16](#)
read_chroms, [17](#)
read_mdf, [19](#)
read_mzml, [20](#)
read_peaklist, [21](#)
read_shimadzu, [22](#)
read_shimadzu_gcd, [23](#)
read_shimadzu_lcd, [24](#)
read_shimadzu_qgd, [26](#)
read_sz_lcd_2d, [28](#)
read_sz_lcd_3d, [29](#)
read_thermoraw, [30](#)
read_varian_peaklist, [32](#)
read_varian_sms, [32](#)
read_waters_arw, [34](#)
read_waters_raw, [35](#)
sp_converter, [36](#)
uv_converter, [36](#)
write_andi_chrom, [37](#)
write_chroms, [38](#)
write_mzml, [39](#)