Package: invitroTKstats (via r-universe)

October 16, 2026

Version 0.0.14

Date 2026-01-22

Title In Vitro Toxicokinetic Data Processing and Analysis Pipeline

Description A set of tools for processing and analyzing in vitro toxicokinetic measurements in a standardized and reproducible pipeline. The package was developed to perform frequentist and Bayesian estimation on a variety of in vitro toxicokinetic measurements including -- but not limited to -- chemical fraction unbound in the presence of plasma (f_up), intrinsic hepatic clearance (Clint, uL/min/million hepatocytes), and membrane permeability for oral absorption (Caco2). The methods provided by the package were described in Wambaugh et al. (2019) <doi:10.1093/toxsci/kfz205>.

Depends R (>= 3.5.0)

Imports runjags, parallel, readxl, coda, ggplot2, scales, stats4, Rdpack, methods, stats, utils, dplyr, rlang

RdMacros Rdpack

Suggests knitr, R.rsp, tidyverse, gridExtra, gridtext, flextable, rmarkdown, magrittr, stringr

License MIT + file LICENSE

LazyData true

Encoding UTF-8

VignetteBuilder knitr, R.rsp

RoxygenNote 7.3.2

NeedsCompilation no

Maintainer Caroline Ring <ring.caroline@epa.gov>

URL https://github.com/USEPA/invitroTKstats

BugReports https://github.com/USEPA/invitroTKstats/issues

Author John Wambaugh [aut] (ORCID:
<https://orcid.org/0000-0002-4024-534X>), Caroline Ring [cre]
(ORCID: <https://orcid.org/0000-0002-0463-1251>), Sarah E.
Davidson-Fritz [aut] (ORCID:
<https://orcid.org/0000-0002-2891-9380>), Lindsay Knupp [ctb],
Barbara A. Wetmore [ctb] (ORCID:
<https://orcid.org/0000-0002-6878-5348>), Zhao Zhihui [ctb],
Chantel Nicolas [ctb] (ORCID:
<https://orcid.org/0000-0003-2061-5830>), Anna Kreutz [ctb]
(ORCID: <https://orcid.org/0000-0002-7787-0683>), U.S. Federal
Government [cph] (Copyright holder of this package)

Repository https://cran.r-universe.dev

Date/Publication 2026-01-26 10:48:45 UTC

RemoteUrl https://github.com/cran/invitroTKstats

RemoteRef HEAD

RemoteSha 9e926839e3a0b2366196b32ef8c0a87a1046270d

Contents

.check_all_miss_cols 4
.check_char_cols 4
.check_no_miss_cols 5
.check_num_cols 5
.check_std_colnames_in_data 6
build_mydata_clint 6
build_mydata_fup_red 7
build_mydata_fup_uc 7
caco2_cheminfo 8
caco2_L0 9
caco2_L1 10
caco2_L2 11
caco2_L3 12
calc_caco2_point 14
calc_clint 17
calc_clint_point 20
calc_fup_red 23
calc_fup_red_point 26
calc_fup_uc 29
calc_fup_uc_point 32
check_catalog 35
clint_cheminfo 36
clint_L0 36
clint_L1 38
clint_L2 39
clint_L2_heldout 40
clint_L3 42

clint_L4 . . . 43
clint_PREJAGS . . . 44
create_catalog . . . 45
create_chem_table . . . 47
create_method_table . . . 49
extract_level1_fup_uc . . . 51
format_caco2 . . . 53
format_clint . . . 60
format_fup_red . . . 66
format_fup_uc . . . 72
fup_red_cheminfo . . . 78
fup_red_L0 . . . 79
fup_red_L1 . . . 80
fup_red_L2 . . . 82
fup_red_L2_heldout . . . 83
fup_red_L3 . . . 84
fup_red_L4 . . . 85
fup_red_PREJAGS . . . 86
fup_uc_cheminfo . . . 88
fup_uc_L0 . . . 88
fup_uc_L1 . . . 89
fup_uc_L2 . . . 91
fup_uc_L2_heldout . . . 92
fup_uc_L3 . . . 93
fup_uc_L4 . . . 94
fup_uc_PREJAGS . . . 95
Heaviside . . . 96
initfunction_clint . . . 96
initfunction_fup_red . . . 97
initfunction_fup_uc . . . 97
L1.common.col . . . 98
merge_level0 . . . 98
plot_clint . . . 103
plot_fup_uc . . . 104
round_output . . . 105
runjagsdata.to.list . . . 108
sample_verification . . . 108
scientific_10 . . . 111
std.catcols . . . 112
summarize_table . . . 112

Index **116**

.check_all_miss_cols *Check if all the data is missing for specified columns.*

Description

This function checks for whether any of the specified columns are missing all of their data, either 'NA' and/or 'NULL'.

Usage

```
.check_all_miss_cols(data, req.cols)
```

Arguments

data	Data frame to check.
req.cols	Column names that should be checked for whether all data is missing.

.check_char_cols *Check the character columns are correctly of character class.*

Description

Check the character columns are correctly of character class.

Usage

```
.check_char_cols(data, char.cols)
```

Arguments

data	Data frame to check.
char.cols	Column names that should be of the character class.

.check_no_miss_cols *Check there is no missing data for specified columns.*

Description

This function checks for whether any of the required columns have a data entry of 'NA' or 'NULL'.

Usage

```
.check_no_miss_cols(data, req.cols, return.missing = FALSE)
```

Arguments

data	Data frame to check.
req.cols	Columns with required data.
return.missing	Logical argument, if 'TRUE' return rows missing data in column (list or vector by column name). (Default is 'FALSE'.)

.check_num_cols *Check the numeric columns are correctly of numeric class.*

Description

Check the numeric columns are correctly of numeric class.

Usage

```
.check_num_cols(data, num.cols)
```

Arguments

data	Data frame to check.
num.cols	Column names that should be of the numeric class.

.check_std_colnames_in_data *Check the standard column names are in the data.*

Description

Check the standard column names are in the data.

Usage

```
.check_std_colnames_in_data(data, std.colnames, data.name = NULL)
```

Arguments

data	Data frame to check.
std.colnames	Vector of character strings with standard column names to check for in the data.
data.name	Name of the data object passed to the standard column names check function. (Defaults to NULL.)

build_mydata_clint *Build Data Object for Intrinsic Hepatic Clearance (Clint) Bayesian Model*

Description

Builds a list of arguments required for JAGS from subset of level-2 data frame. The list is used as an argument to JAGS during level-4 processing.

Usage

```
build_mydata_clint(
  this.cvt,
  this.data,
  decrease.prob,
  saturate.prob,
  degrade.prob
)
```

Arguments

this.cvt	(Data Frame) Subset of data containing all "Cvst" sample observations of one test compound.
this.data	(Data Frame) Subset of data containing all observations of one test compound.
decrease.prob	(Numeric) Prior probability that a chemical will decrease in the assay.

saturate.prob (Numeric) Prior probability that a chemicals rate of metabolism will decrease between 1 and 10 uM.

degrade.prob (Numeric) Prior probability that a chemical will be unstable (that is, degrade abiotically) in the assay.

Value

A named list to be passed into the Bayesian model.

`build_mydata_fup_red` *Build Data Object for Fup RED Bayesian Model*

Description

Builds a list of arguments required for JAGS from subset of level-2 data frame. The list is used as an argument to JAGS during level-4 processing.

Usage

```
build_mydata_fup_red(this.data, Physiological.Protein.Conc)
```

Arguments

this.data (Data Frame) Subset of data containing all observations of one test compound.

Physiological.Protein.Conc
(Numeric) The assumed physiological protein concentration for plasma protein binding calculations.

Value

A named list to be passed into the Bayesian model.

`build_mydata_fup_uc` *Build Data Object for Fup UC Bayesian Model*

Description

Builds a list of arguments required for JAGS from subset of level-2 data frame. The list is used as an argument to JAGS during level-4 processing.

Usage

```
build_mydata_fup_uc(MS.data, CC.data, T1.data, T5.data, AF.data)
```

Arguments

MS.data	(Data Frame) Subset of data containing all observations of one test compound.
CC.data	(Data Frame) Subset of data containing observations of calibration curves samples.
T1.data	(Data Frame) Subset of data containing observations of Whole Plasma T1h Samples.
T5.data	(Data Frame) Subset of data containing observations of Whole Plasma T5h Samples.
AF.data	(Data Frame) Subset of data containing observations of Aqueous Fraction samples.

Value

A named list to be passed into the Bayesian model.

---

caco2_cheminfo *Caco-2 Chemical Information Example Data set*

Description

The chemical ID mapping information from tandem mass spectrometry (MS/MS) measurements of Caco-2 assay-specific data (Honda et al. 2025) . This data set contains 520 unique compounds/chemicals.

Usage

```
caco2_cheminfo
```

Format

A chemical info data.frame with 554 rows and 7 variables:

DTXSID DSSTox Substance Identifier (CompTox Chemicals Dashboard)

PREFERRED_NAME Preferred compound name from the CompTox Chemicals Dashboard (CCD)

CASRN CAS Registry Number of the test compound

MOLECULAR_FORMULA Molecular formula of the test compound

AVERAGE_MASS Molecular weight of the compound in daltons

QSAR_READY_SMILES SMILES (Simplified molecular-input line-entry system) chemical structure description.

test_article Compound ID used in the laboratory

References

Honda GS, Kenyon EM, Davidson-Fritz S, Dinallo R, El Masri H, Korol-Bexell E, Li L, Angus D, Pearce RG, Sayre RR, others (2025). "Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment." *ALTEX-Alternatives to animal experimentation*, **42**(1), 56–74.

caco2_L0 *Caco-2 Level-0 Example Data set*

Description

A subset of tandem mass spectrometry (MS/MS) measurements of Caco-2 assay-specific data (Honda et al. 2025). This subset contains samples for 3 test analytes/compounds.

Usage

```
caco2_L0
```

Format

A level-0 data.frame with 48 rows and 17 variables:

`Compound` Compound name

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard)

`Lab.Compound.ID` Compound ID used in the laboratory

`Date` Date MS/MS assay data acquired from instrument

`Sample` Sample Name

`Type` Type of Caco-2 sample

`Compound.Conc` Expected (or nominal) concentration of analyte (for calibration curve)

`Peak.Area` Peak area of analyte (target compound)

`ISTD.Peak.Area` Peak area of internal standard (pixels)

`ISTD.Name` Name of compound used as internal standard (ISTD)

`Analysis.Params` General description of chemical analysis method

`Level0.File` Name of data file from laboratory that was used to compile level-0 data.frame

`Level0.Sheet` Name of "sheet" (for Excel workbooks) from which the laboratory data were read

`Direction` Direction of the Caco-2 permeability experiment

`Vol.Donor` The media volume (in cm^3) of the donor portion of the Caco-2 experimental well

`Vol.Receiver` The media volume (in cm^3) of the receiver portion of the Caco-2 experimental well

`Dilution.Factor` Number of times the sample was diluted

References

Honda GS, Kenyon EM, Davidson-Fritz S, Dinallo R, El Masri H, Korol-Bexell E, Li L, Angus D, Pearce RG, Sayre RR, others (2025). "Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment." *ALTEX-Alternatives to animal experimentation*, **42**(1), 56–74.

caco2_L1 *Caco-2 Level-1 Example Data set*

Description

A subset of tandem mass spectrometry (MS/MS) measurements of Caco-2 assay-specific data (Honda et al. 2025). This subset contains samples for 3 test analytes/compounds.

Usage

```
caco2_L1
```

Format

A level-1 data.frame with 48 rows and 28 variables:

`Lab.Sample.Name` Sample name as described in the laboratory

`Date` Date MS/MS assay data acquired from instrument

`Compound.Name` Compound name

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of Caco-2 sample

`Direction` Direction of the Caco-2 permeability experiment

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to donor side (uM)

`Time` Time when sample was measured (h)

`ISTD.Name` Name of compound used as internal standard (ISTD)

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (pixels)

`Area` Peak area of analyte (target compound)

`Membrane.Area` The area of the Caco-2 monolayer.

`Vol.Donor` The media volume (in cm^3) of the donor portion of the Caco-2 experimental well

`Vol.Receiver` The media volume (in cm^3) of the receiver portion of the Caco-2 experimental well

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Additional information

`Level0.File` Name of data file from laboratory that was used to compile level-0 data.frame)

`Level0.Sheet` Name of "sheet" (for Excel workbooks) from which the laboratory data were read

`Response` Response factor (calculated from analyte and ISTD peaks)

References

Honda GS, Kenyon EM, Davidson-Fritz S, Dinallo R, El Masri H, Korol-Bexell E, Li L, Angus D, Pearce RG, Sayre RR, others (2025). "Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment." *ALTEX-Alternatives to animal experimentation*, **42**(1), 56–74.

`caco2_L2` *Caco-2 Level-2 Example Data set*

Description

A subset of tandem mass spectrometry (MS/MS) measurements of Caco-2 assay-specific data (Honda et al. 2025). This subset contains samples for 3 test analytes/compounds.

Usage

```
caco2_L2
```

Format

A level-2 data.frame with 48 rows and 29 variables:

`Lab.Sample.Name` Sample name as described in the laboratory

`Date` Date MS/MS assay data acquired from instrument

`Compound.Name` Compound name

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of Caco-2 sample

`Direction` Direction of the Caco-2 permeability experiment

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to donor side (uM)

`Time` Time when sample was measured (h)

`ISTD.Name` Name of compound used as internal standard (ISTD)

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (pixels)

`Area` Peak area of analyte (target compound)

`Membrane.Area` The area of the Caco-2 monolayer.

`Vol.Donor` The media volume (in cm^3) of the donor portion of the Caco-2 experimental well

`Vol.Receiver` The media volume (in cm^3) of the receiver portion of the Caco-2 experimental well

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Additional information

`Level0.File` Name of data file from laboratory that was used to compile level-0 data.frame)

`Level0.Sheet` Name of "sheet" (for Excel workbooks) from which the laboratory data were read

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other causes the data to be ignored.)

References

Honda GS, Kenyon EM, Davidson-Fritz S, Dinallo R, El Masri H, Korol-Bexell E, Li L, Angus D, Pearce RG, Sayre RR, others (2025). "Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment." *ALTEX-Alternatives to animal experimentation*, **42**(1), 56–74.

`caco2_L3` *Caco-2 Level-3 Example Data set*

Description

A subset of tandem mass spectrometry (MS/MS) measurements of Caco-2 assay-specific data (Honda et al. 2025). This subset contains samples for 3 test analytes/compounds.

Usage

```
caco2_L3
```

Format

A level-3 data.frame with 3 rows and 20 variables:

`Compound.Name` Compound name

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard)

`Time` Time when sample was measured (h)

`Membrane.Area` The area of the Caco-2 monolayer

`Calibration` Identifier for mass spectrometry calibration – usually the date

`C0_A2B` Initial concentration in the apical side

`dQdt_A2B` Rate of permeation from the apical to the basolateral side

`Papp_A2B` Apparent membrane permeability from the apical to the basolateral side

`Frec_A2B.vec` Fraction of the initial compound in the apical side recovered in the basolateral side (collapsed numeric vector, values for replicates separated by a "|")

`Frec_A2B.mean` Mean of fraction recovered values in the apical to basolateral direction

`Recovery_Class_A2B.vec` Recovery classification of fraction recovered values in the apical to basolateral direction (collapsed character vector, values for replicates separated by a "|")

`Recovery_Class_A2B.mean` Recovery classification of mean fraction recovered in the apical to basolateral direction

`C0_B2A` Initial concentration in the basolateral side

`dQdt_B2A` Rate of permeation from the basolateral to the apical side

`Papp_B2A` Apparent membrane permeability from the basolateral to the apical side

`Frec_B2A.vec` Fraction of the initial compound in the basolateral side recovered in the apical side (collapsed numeric vector, values for replicates separated by a "|")

`Frec_B2A.mean` Mean of fraction recovered values in the basolateral to apical direction

`Recovery_Class_B2A.vec` Recovery classification of fraction recovered values in the basolateral to apical direction (collapsed character vector, values for replicates separated by a "|")

`Recovery_Class_B2A.mean` Recovery classification of mean fraction recovered in the basolateral to apical direction

`Refflux` Efflux ratio

References

Honda GS, Kenyon EM, Davidson-Fritz S, Dinallo R, El Masri H, Korol-Bexell E, Li L, Angus D, Pearce RG, Sayre RR, others (2025). "Impact of gut permeability on estimation of oral bioavailability for chemicals in commerce and the environment." *ALTEX-Alternatives to animal experimentation*, **42**(1), 56–74.

`calc_caco2_point` *Calculate a Point Estimate of Apparent Membrane Permeability (Papp) from Caco-2 data (Level-3)*

Description

This function calculates a point estimate of apparent membrane permeability (Papp) using mass spectrometry (MS) peak areas from samples collected as part of in vitro measurements of membrane permeability using Caco-2 cells (Hubatsch et al. 2007).

Usage

```
calc_caco2_point(
  FILENAME,
  data.in,
  good.col = "Verified",
  output.res = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

`FILENAME`	(Character) A string used to identify the input level-2 file, "<FILENAME>-Caco-2-Level2.tsv" (if importing from a .tsv file), and/or used to identify the output level-3 file, "<FILENAME>-Caco-2-Level3.tsv" (if exporting).
`data.in`	(Data Frame) A level-2 data frame generated from the `format_caco2` function with a verification column added by `sample_verification`. Complement with manual verification if needed.
`good.col`	(Character) Column name indicating which rows have been verified, data rows valid for analysis are indicated with a "Y". (Defaults to "Verified".)
`output.res`	(Logical) When set to `TRUE`, the result table (level-3) will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)
`sig.figs`	(Numeric) The number of significant figures to round the exported result table (level-3). (Note: console print statements are also rounded to specified significant figures.) (Defaults to `3`.)
`INPUT.DIR`	(Character) Path to the directory where the input level-2 file exists. If `NULL`, looking for the input level-2 file in the current working directory. (Defaults to `NULL`.)
`OUTPUT.DIR`	(Character) Path to the directory to save the output file. If `NULL`, the output file will be saved to the user's per-session temporary directory or `INPUT.DIR` if specified. (Defaults to `NULL`.)
`verbose`	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the `format_caco2` function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

The data frame of observations should be annotated according to direction (either apical to basolateral – "AtoB" – or basolateral to apical – "BtoA") and type of concentration measured:

Blank with no chemical added	Blank
Target concentration added to donor compartment at time 0 (C0)	D0
Donor compartment at end of experiment	D2
Receiver compartment at end of experiment	R2

Apparent membrane permeability (P_{app}) is calculated from MS responses as:

$P_{app} = \frac{dQ/dt}{c_0 * A}$

The rate of permeation, $\frac{dQ}{dt}\left(\frac{\text{peak area}}{\text{time (s)}}\right)$ is calculated as:

$\frac{dQ}{dt} = \max\left(0, \frac{\sum_{i=1}^{n_{R2}}(r_{R2} * c_{DF})}{n_{R2}} - \frac{\sum_{i=1}^{n_{BL}}(r_{BL} * c_{DF})}{n_{BL}}\right)$

where r_{R2} is Receiver Response, c_{DF} is the corresponding Dilution Factor, r_{BL} is Blank Response, n_{R2} is the number of Receiver Responses, and n_{BL} is the number of Blank Responses.

If the output level-3 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

`data.frame` A level-3 data.frame in standardized format

C0_A2B	
dQdt_A2B	
Papp_A2B	
C0_B2A	
dQdt_B2A	
Papp_B2A	
Refflux	
Frec_A2B.vec	Fraction recovered for the apical-basolateral direction, calculated as the fraction of the initial d
Frec_A2B.mean	
Frec_B2A.vec	Fraction recovered for the basolateral-apical direction, calculated in the same way as F
Frec_B2A.mean	
Recovery_Class_A2B.vec	Recovery classification for apical-to-basolateral permeability("Low Recovery" if Frec_A2B.vec
Recovery_Class_A2B.mean	Recovery classification for the mean apic

Recovery_Class_B2A.vec Recovery classification for basolateral-to-apical permeability("Low Recovery" if Frec_B2A.vec
Recovery_Class_B2A.mean Recovery classification for the mean basc

Author(s)

John Wambaugh

References

Hubatsch I, Ragnarsson EG, Artursson P (2007). "Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers." *Nature protocols*, **2**(9), 2111–2119.

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::caco2_L2

## scenario 1:
## input level-2 data from the R session and do not export the result table
level3 <- calc_caco2_point(data.in = level2, output.res = FALSE)

## scenario 2:
## import level-2 data from a 'tsv' file and export the result table to
## same location as INPUT.DIR
## Not run:
## Refer to sample_verification help file for how to export level-2 data to a directory.
## Unless a different path is specified in OUTPUT.DIR,
## the result table will be saved to the directory specified in INPUT.DIR.
## Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
level3 <- calc_caco2_point(# e.g. replace with "Examples" from "Examples-Caco-2-Level2.tsv"
                           FILENAME="<level-2 FILENAME prefix>",
                           INPUT.DIR = "<level-2 FILE LOCATION>",
                           output.res = TRUE)

## End(Not run)

## scenario 3:
## input level-2 data from the R session and export the result table to the
## user's temporary directory
## Will need to replace FILENAME with desired level-2 filename prefix.
## Not run:
level3 <- calc_caco2_point(# e.g. replace with "MYDATA"
                           FILENAME = "<desired level-2 FILENAME prefix>",
                           data.in = level2,
                           output.res = TRUE)
# To delete, use the following code. For more details, see the link in the
# "Details" section.
file.remove(list.files(tempdir(), full.names = TRUE,
pattern = "<desired level-2 FILENAME prefix>-Caco-2-Level3.tsv"))

## End(Not run)
```

`calc_clint` *Calculate Intrinsic Hepatic Clearance (Clint) with Bayesian Modeling (Level-4)*

Description

This function estimates the intrinsic hepatic clearance (Clint) with Bayesian modeling on Hepatocyte Incubation data (Shibata et al. 2002). Clint and the credible intervals, at both 1 and 10 uM (if tested), are estimated from posterior samples of the MCMC. A summary table (level-4) along with the full set of MCMC results is returned from the function.

Usage

```
calc_clint(
  FILENAME,
  data.in,
  TEMP.DIR = NULL,
  NUM.CHAINS = 5,
  NUM.CORES = 2,
  RANDOM.SEED = 1111,
  SEED.SET = NULL,
  good.col = "Verified",
  JAGS.PATH = NA,
  decrease.prob = 0.5,
  saturate.prob = 0.25,
  degrade.prob = 0.05,
  save.MCMC = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

`FILENAME`	(Character) A string used to identify the input level-2 file, "<FILENAME>-Clint-Level2.tsv", and to name the exported model results. This argument is required no matter which method of specifying input data is used. (Defaults to `NULL`.)
`data.in`	(Data Frame) A level-2 data frame generated from the `format_clint` function with a verification column added by `sample_verification`. Complement with manual verification if needed.
`TEMP.DIR`	(Character) Temporary directory to save intermediate files. If `NULL`, all files will be written to the user's per-session temporary directory. (Defaults to `NULL`.)

NUM.CHAINS	(Numeric) The number of Markov Chains to use. (Defaults to 5.)
NUM.CORES	(Numeric) The number of processors to use for parallel computing. (Defaults to 2.)
RANDOM.SEED	(Numeric) The seed used by the random number generator. (Defaults to 1111.)
SEED.SET	(Numeric Vector) A set of seeds used by the random number generator for each chain. Should be unique for each chain and vector length should equal the total number of chains. (Default is NULL.)
good.col	(Character) Column name indicating which rows have been verified for analysis, valid data rows are indicated with "Y". (Defaults to "Verified".)
JAGS.PATH	(Character) Computer specific file path to JAGS software. (Defaults to NA.)
decrease.prob	(Numeric) Prior probability that a chemical will decrease in the assay. (Defaults to 0.5.)
saturate.prob	(Numeric) Prior probability that a chemicals rate of metabolism will decrease between 1 and 10 uM. (Defaults to 0.25.)
degrade.prob	(Numeric) Prior probability that a chemical will be unstable (that is, degrade abiotically) in the assay. (defaults to 0.05.)
save.MCMC	(Logical) When set to TRUE, will export the MCMC results as an .RData file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported unverified data (level-2). The exported result table (level-4) is left unrounded for reproducibility. (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the format_clint function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

Note: By default, this function writes files to the user's per-session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

Users must specify an alternative path with the TEMP.DIR argument if they want the intermediate files exported to another path. Exported intermediate files include the summary results table (.tsv), JAGS model (.RData), and any "unverified" data excluded from the analysis (.tsv). Users must specify an alternative path with the OUTPUT.DIR argument if they want the final output file exported to another path. The exported final output file is the summary results table (.RData).

As a best practice, INPUT.DIR (when importing a .tsv file) and/or OUTPUT.DIR should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

The data frame of observations should be annotated according to these types:

Blank	Cell free blank with media
CC	Cell free calibration curve
Cvst	Hepatocyte incubation concentration vs. time
Inactive	Concentration vs. time data with inactivated hepatocytes

We currently require Cvst data. Blank, CC, and Inactive data are optional.

Clint is calculated using lm to perform a linear regression of MS response as a function of time.

Additional User Notification(s):

- runjags::findjags() may not work as JAGS.PATH argument. Instead, may need to manually remove the trailing path such that JAGS.PATH only contains path information through "/x64" (e.g. JAGS.PATH = "/Program Files/JAGS/JAGS-4.3.1/x64").

Value

A list of two objects:

1. Results: A level-4 data frame with the Bayesian estimated intrinsic hepatic clearance (Clint) for 1 and 10 uM and credible intervals for all compounds in the input file. Column includes: Compound.Name - compound name, Lab.Compound.Name - compound name used by the laboratory, DTXSID - EPA's DSSTox Structure ID, Clint.1.Med/Clint.10.Med - posterior median, Clint.1.Low/Clint.10.Low - 2.5th quantile, Clint.1.High/Clint.10.High - 97.5th quantile, Clint.pValue, Sat.pValue, degrades.pValue - "p-values" estimated from the probabilities of observing decreases, saturations, and abiotic degradations in all posterior samples.
2. coda: A runjags-class object containing results from JAGS model.

Author(s)

John Wambaugh

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Examples

```
## Example 1: loading level-2 using data.in and export all files to the user's
## temporary directory
## Not run:
level2 <- invitroTKstats::clint_L2
```

```
# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_clint(FILENAME = "Example1",
                     data.in = level2,
                     NUM.CORES=2,
                     JAGS.PATH=path.to.JAGS)

## End(Not run)

## Example 2: importing level-2 from a .tsv file and export all files to same
## location as INPUT.DIR
## Not run:
# Refer to sample_verification help file for how to export level-2 data to a directory.
# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
# Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_clint(# e.g. replace with "Examples" from "Examples-Clint-Level2.tsv"
                     FILENAME="<level-2 FILENAME prefix>",
                     NUM.CORES=2,
                     JAGS.PATH=path.to.JAGS,
                     INPUT.DIR = "<level-2 FILE LOCATION>")

## End(Not run)
```

`calc_clint_point` *Calculate a Point Estimate of Intrinsic Hepatic Clearance (Clint) (Level-3)*

Description

This function calculates a point estimate of intrinsic hepatic clearance (Clint) using mass spectrometry (MS) peak area data collected as part of *in vitro* measurements of chemical clearance, as characterized by the disappearance of parent compound over time when incubated with primary hepatocytes (Shibata et al. 2002).

Usage

```
calc_clint_point(
  FILENAME,
  data.in,
  good.col = "Verified",
  output.res = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
```

```
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	A string used to identify the input level-2 file, "<FILENAME>-Clint-Level2.tsv" (if importing from a .tsv file), and/or used to identify the output level-3 file, "<FILENAME>-Clint-Level3.tsv" (if exporting).
data.in	(Data Frame) A level-2 data frame generated from the format_clint function with a verification column added by sample_verification. Complement with manual verification if needed.
good.col	(Character) Column name indicating which rows have been verified, data rows valid for analysis are indicated with a "Y". (Defaults to "Verified".)
output.res	(Logical) When set to TRUE, the result table (level-3) will be exported to the user's per-session temporary directory or OUTPUT.DIR (if specified) as a .tsv file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported result table (level-3). (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the format_clint function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

The data frame of observations should be annotated according to these types:

	Blank	Blank
Hepatocyte incubation concentration vs. time		Cvst

Clint is calculated using lm to perform a linear regression of MS response as a function of time.

If the output level-3 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: tempdir(). For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, INPUT.DIR (when importing a .tsv file) and/or OUTPUT.DIR should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-3 data frame with one row per chemical, contains a point estimate of intrinsic clearance (Clint), estimates of Clint of assays performed at 1 and 10 uM (if tested), the p-value and the Akaike Information Criterion (AIC) of the linear regression fit for all chemicals in the input data frame.

Author(s)

John Wambaugh

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::clint_L2

## scenario 1:
## input level-2 data from the R session and do not export the result table
level3 <- calc_clint_point(data.in = level2, output.res = FALSE)

## scenario 2:
## import level-2 data from a 'tsv' file and export the result table to
## same location as INPUT.DIR
## Not run:
## Refer to sample_verification help file for how to export level-2 data to a directory.
## Unless a different path is specified in OUTPUT.DIR,
## the result table will be saved to the directory specified in INPUT.DIR.
## Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
level3 <- calc_clint_point(# e.g. replace with "Examples" from "Examples-Clint-Level2.tsv"
                           FILENAME="<level-2 FILENAME prefix>",
                           INPUT.DIR = "<level-2 FILE LOCATION>",
                           output.res = TRUE)

## End(Not run)

## scenario 3:
## input level-2 data from the R session and export the result table to the
## user's temporary directory
## Will need to replace FILENAME with desired level-2 filename prefix.
## Not run:
level3 <- calc_clint_point(# e.g. replace with "MYDATA"
                           FILENAME = "<desired level-2 FILENAME prefix>",
                           data.in = level2,
                           output.res = TRUE)
# To delete, use the following code. For more details, see the link in the
# "Details" section.
```

```
file.remove(list.files(tempdir(), full.names = TRUE,
pattern = "<desired level-2 FILENAME prefix>-Clint-Level3.tsv"))

## End(Not run)
```

`calc_fup_red` *Calculate Fraction Unbound in Plasma (Fup) from Rapid Equilibrium Dialysis (RED) Data with Bayesian Modeling (Level-4)*

Description

This function estimates the fraction unbound in plasma (Fup) with Bayesian modeling on Rapid Equilibrium Dialysis (RED) data (Waters et al. 2008). Both Fup and the credible interval are estimated from posterior samples of the MCMC. A summary table (level-4) along with the full set of MCMC results is returned from the function.

Usage

```
calc_fup_red(
  FILENAME,
  data.in,
  TEMP.DIR = NULL,
  NUM.CHAINS = 5,
  NUM.CORES = 2,
  RANDOM.SEED = 1111,
  SEED.SET = NULL,
  good.col = "Verified",
  JAGS.PATH = NA,
  Physiological.Protein.Conc = 70/(66.5 * 1000) * 1e+06,
  save.MCMC = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

`FILENAME`	(Character) A string used to identify the input level-2 file, "<FILENAME>-fup-RED-Level2.tsv", and to name the exported model results. This argument is required no matter which method of specifying input data is used. (Defaults to `NULL`.)
`data.in`	(Data Frame) A level-2 data frame generated from the `format_fup_red` function with a verification column added by `sample_verification`. Complement with manual verification if needed.

TEMP.DIR	(Character) Temporary directory to save intermediate files. If NULL, all files will be written to the user's per-session temporary directory. (Defaults to NULL.)
NUM.CHAINS	(Numeric) The number of Markov Chains to use. (Defaults to 5.)
NUM.CORES	(Numeric) The number of processors to use for parallel computing. (Defaults to 2.)
RANDOM.SEED	The seed used by the random number generator. (Defaults to 1111.)
SEED.SET	(Numeric Vector) A set of seeds used by the random number generator for each chain. Should be unique for each chain and vector length should equal the total number of chains. (Default is NULL.)
good.col	(Character) Column name indicating which rows have been verified for analysis, valid data rows are indicated with "Y". (Defaults to "Verified".)
JAGS.PATH	(Character) Computer specific file path to JAGS software. (Defaults to NA.)
Physiological.Protein.Conc	(Numeric) The assumed physiological protein concentration for plasma protein binding calculations. (Defaults to 70/(66.5*1000)*1000000. According to Berg and Lane (2011): 60-80 mg/mL, albumin is 66.5 kDa, assume all protein is albumin to estimate default in uM.)
save.MCMC	(Logical) When set to TRUE, will export the MCMC results as an .RData file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported unverified data (level-2). The exported result table (level-4) is left unrounded for reproducibility. (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1 data, formatted with the format_fup_red function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

Note: By default, this function writes files to the user's per-session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

Users must specify an alternative path with the TEMP.DIR argument if they want the intermediate files exported to another path. Exported intermediate files include the summary results table (.tsv), JAGS model (.RData), and any "unverified" data excluded from the analysis (.tsv). Users must specify an alternative path with the OUTPUT.DIR argument if they want the final output file exported to another path. The exported final output file is the summary results table (.RData).

As a best practice, INPUT.DIR (when importing a .tsv file) and/or OUTPUT.DIR should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

The data frame of observations should be annotated according to of these types:

No Plasma Blank (no chemical, no plasma)	NoPlasma.Blank
Plasma Blank (no chemical, just plasma)	Plasma.Blank
Time zero chemical and plasma	T0
Equilibrium chemical in phosphate-buffered well (no plasma)	PBS
Equilibrium chemical in plasma well	Plasma
Calibration Curve	CC

We currently require Plasma, PBS, and Plasma.Blank data. T0, CC, and NoPlasma.Blank data are optional.

Additional User Notification(s):

- runjags::findjags() may not work as JAGS.PATH argument. Instead, may need to manually remove the trailing path such that JAGS.PATH only contains path information through "/x64" (e.g. JAGS.PATH = "/Program Files/JAGS/JAGS-4.3.1/x64").

Value

A list of two objects:

1. Results: A level-4 data frame with the Bayesian estimated fraction unbound in plasma (Fup) and credible interval for all compounds in the input file. Column includes: Compound.Name - compound name, Lab.Compound.Name - compound name used by the laboratory, DTXSID - EPA's DSSTox Structure ID, Fup.point - point estimate of Fup, Fup.Med - posterior median, Fup.Low - 2.5th quantile, and Fup.High - 97.5th quantile
2. coda: A runjags-class object containing results from JAGS model.

Author(s)

John Wambaugh and Chantel Nicolas

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Wambaugh JF, Wetmore BA, Ring CL, Nicolas CI, Pearce RG, Honda GS, Dinallo R, Angus D, Gilbert J, Sierra T, others (2019). "Assessing toxicokinetic uncertainty and variability in risk prioritization." *Toxicological Sciences*, **172**(2), 235–251.

Berg J, Lane V (2011). "Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory." *Annals of Clinical Biochemistry*, **48**(3), 195–197.

Examples

```
## Example 1: loading level-2 using data.in and export all files to the user's
## temporary directory
## Not run:
level2 <- invitroTKstats::fup_red_L2

# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_fup_red(FILENAME = "Example1",
                       data.in = level2,
                       NUM.CORES=2,
                       JAGS.PATH=path.to.JAGS)

## End(Not run)

## Example 2: importing level-2 from a .tsv file and export all files to same
## location as INPUT.DIR
## Not run:
# Refer to sample_verification help file for how to export level-2 data to a directory.
# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
# Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_fup_red(# e.g. replace with "Examples" from "Examples-fup-RED-Level2.tsv"
                       FILENAME="<level-2 FILENAME prefix>",
                       NUM.CORES=2,
                       JAGS.PATH=path.to.JAGS,
                       INPUT.DIR = "<level-2 FILE LOCATION>")

## End(Not run)
```

`calc_fup_red_point` *Calculate Point Estimates of Fraction Unbound in Plasma (Fup) with Rapid Equilibrium Dialysis (RED) Data (Level-3)*

Description

This function calculates the point estimates for the fraction unbound in plasma (Fup) using mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of chemical Fup using rapid equilibrium dialysis (Waters et al. 2008). See the Details section for the equation(s) used in point estimation.

Usage

```
calc_fup_red_point(
```

```
  FILENAME,
  data.in,
  good.col = "Verified",
  output.res = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the input level-2 file, "<FILENAME>-fup-RED-Level2.tsv" (if importing from a .tsv file), and/or used to identify the output level-3 file, "<FILENAME>-fup-RED-Level3.tsv" (if exporting).
data.in	(Data Frame) A level-2 data frame generated from the format_fup_red function with a verification column added by sample_verification. Complement with manual verification if needed.
good.col	(Character) Column name indicating which rows have been verified, data rows valid for analysis are indicated with a "Y". (Defaults to "Verified".)
output.res	(Logical) When set to TRUE, the result table (level-3) will be exported to the user's per-session temporary directory or OUTPUT.DIR (if specified) as a .tsv file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported result table (level-3). (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the format_fup_red function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

The data frame of observations should be annotated according to these types:

No Plasma Blank (no chemical, no plasma)	NoPlasma.Blank
Plasma Blank (no chemical, just plasma)	Plasma.Blank
Time zero chemical and plasma	T0
Equilibrium chemical in phosphate-buffered well (no plasma)	PBS
Equilibrium chemical in plasma well	Plasma

f_{up} is calculated from MS responses as:

$$f_{up} = \frac{\max\left(0, \frac{\sum_{i=1}^{n_P}(r_{P} * c_{DF})}{n_P} - \frac{\sum_{i=1}^{n_{NPB}}(r_{NPB} * c_{DF})}{n_{NPB}}\right)}{\frac{\sum_{i=1}^{n_{PL}}(r_{PL} * c_{DF})}{n_{PL}} - \frac{\sum_{i=1}^{n_B}(r_B * c_{DF})}{n_B}}$$

where r_P is PBS Response, n_P is the number of PBS Responses, c_{DF} is the corresponding Dilution Factor, r_{NPB} is No Plasma Blank Response, n_{NPB} is the number of No Plasma Blank Responses, r_{PL} is Plasma Response, n_{PL} is the number of Plasma Responses, r_B is Plasma Blank Response, and n_B is the number of Plasma Blank Responses.

If the output level-3 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-3 data frame with one row per chemical, contains chemical identifiers such as preferred compound name, EPA's DSSTox Structure ID, calibration details, and point estimates for the fraction unbound in plasma (Fup) for all chemicals in the input data frame.

Author(s)

John Wambaugh

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::fup_red_L2

## scenario 1:
## input level-2 data from the R session and do not export the result table
level3 <- calc_fup_red_point(data.in = level2, output.res = FALSE)

## scenario 2:
## import level-2 data from a 'tsv' file and export the result table
## Not run:
## Refer to sample_verification help file for how to export level-2 data to a directory.
## Unless a different path is specified in OUTPUT.DIR,
## the result table will be saved to the directory specified in INPUT.DIR.
```

```
## Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
level3 <- calc_fup_red_point(# e.g. replace with "Examples" from "Examples-fup-RED-Level2.tsv"
                             FILENAME="<level-2 FILENAME prefix>",
                             INPUT.DIR = "<level-2 FILE LOCATION>",
                             output.res = TRUE)

## End(Not run)

## scenario 3:
## import level-2 data from the R session and export the result table to the
## user's temporary directory
## Will need to replace FILENAME with desired level-2 filename prefix.
## Not run:
level3 <- calc_fup_red_point(# e.g. replace with "MYDATA",
                             FILENAME = "<desired level-2 FILENAME prefix>",
                             data.in = level2,
                             output.res = TRUE)
# To delete, use the following code. For more details, see the link in the
file.remove(list.files(tempdir(), full.names = TRUE,
pattern = "<desired level-2 FILENAME prefix>-fup-RED-Level3.tsv"))

## End(Not run)
```

`calc_fup_uc` *Calculate Fraction Unbound in Plasma (Fup) from Ultracentrifugation (UC) Data with Bayesian Modeling (Level-4)*

Description

This function estimates the fraction unbound in plasma (Fup) and credible intervals with a Bayesian modeling approach, via MCMC simulations. Data used in modeling is collected from Ultracentrifugation (UC) Fup assays (Redgrave et al. 1975). Fup and the credible interval are calculated from the MCMC posterior samples and the function returns a summary table (level-4) along with the full set of MCMC results.

Usage

```
calc_fup_uc(
  FILENAME,
  data.in,
  TEMP.DIR = NULL,
  NUM.CHAINS = 5,
  NUM.CORES = 2,
  RANDOM.SEED = 1111,
  SEED.SET = NULL,
  good.col = "Verified",
  JAGS.PATH = NA,
  save.MCMC = FALSE,
```

```
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the input level-2 file, "<FILENAME>-fup-UC-Level2.tsv", and to name the exported model results. This argument is required no matter which method of specifying input data is used. (Defaults to NULL.)
data.in	A level-2 data frame generated from the format_fup_uc function with a verification column added by sample_verification. Complement with manual verification if needed.
TEMP.DIR	(Character) Temporary directory to save intermediate files. If NULL, all files will be written to the user's per-session temporary directory. (Defaults to NULL.)
NUM.CHAINS	(Numeric) The number of Markov Chains to use. (Defaults to 5.)
NUM.CORES	(Numeric) The number of processors to use for parallel computing. (Defaults to 2.)
RANDOM.SEED	(Numeric) The seed used by the random number generator. (Defaults to 1111.)
SEED.SET	(Numeric Vector) A set of seeds used by the random number generator for each chain. Should be unique for each chain and vector length should equal the total number of chains. (Default is NULL.)
good.col	(Character) Column name indicating which rows have been verified for analysis, valid data rows are indicated with "Y". (Defaults to "Verified".)
JAGS.PATH	(Character) Computer specific file path to JAGS software. (Defaults to 'NA'.)
save.MCMC	(Logical) When set to TRUE, will export the MCMC results as an .RData file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported unverified data (level-2). The exported result table (level-4) is left unrounded for reproducibility. (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the format_fup_uc function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

Note: By default, this function writes files to the user's per-session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

Users must specify an alternative path with the `TEMP.DIR` argument if they want the intermediate files exported to another path. Exported intermediate files include the summary results table (.tsv), JAGS model (.RData), and any "unverified" data excluded from the analysis (.tsv). Users must specify an alternative path with the `OUTPUT.DIR` argument if they want the final output file exported to another path. The exportef final output file is the summary results table (.RData).

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

The data frame of observations should be annotated according to these types:

Calibration Curve	CC
Ultracentrifugation Aqueous Fraction	AF
Whole Plasma T1h Sample	T1
Whole Plasma T5h Sample	T5

We currently require CC, AF, and T5 data. T1 data are optional.

Additional User Notification(s):

- runjags::findjags() may not work as `JAGS.PATH` argument. Instead, may need to manually remove the trailing path such that `JAGS.PATH` only contains path information through "/x64" (e.g. `JAGS.PATH` = "/Program Files/JAGS/JAGS-4.3.1/x64").

Value

A list of two objects:

1. Results: A level-4 data frame with Bayesian estimated fraction unbound in plasma (Fup) and credible intervals for all compounds in the input file. Column includes: Compound.Name - compound name, Lab.Compound.Name - compound name used by the laboratory, DTXSID - EPA's DSSTox Structure ID, Fup.point - point estimate of Fup, Fup.Med - posterior median, Fup.Low - 2.5th quantile, Fup.High - 97.5th quantile, Fstable.Med - posterior median of stability fraction, Fstable.Low - 2.5th quantile, Fstable.High - 97.5th quantile.
2. coda: A runjags-class object containing results from JAGS model.

Author(s)

John Wambaugh and Chantel Nicolas

References

Redgrave TG, Roberts DCK, West CE (1975). "Separation of plasma lipoproteins by density-gradient ultracentrifugation." *Analytical Biochemistry*, **65**(1–2), 42–49.

Examples

```
## Example 1: loading level-2 using data.in and export all files to the user's
## temporary directory
## Not run:
level2 <- invitroTKstats::fup_uc_L2

# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_fup_uc(FILENAME = "Example1",
                      data.in = level2,
                      NUM.CORES=2,
                      JAGS.PATH=path.to.JAGS)

## End(Not run)

## Example 2: importing level-2 from a .tsv file and export all files to same
## location as INPUT.DIR
## Not run:
# Refer to sample_verification help file for how to export level-2 data to a directory.
# JAGS.PATH should be changed to user's specific computer file path to JAGS software.
# findJAGS() from runjags package is a handy function to find JAGS path automatically.
# In certain circumstances or cases, one may need to provide the absolute path to JAGS.
# Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
path.to.JAGS <- runjags::findJAGS()
level4 <- calc_fup_uc(# e.g. replace with "Examples" from "Examples-fup-UC-Level2.tsv"
                      FILENAME="<level-2 FILENAME prefix>",
                      NUM.CORES=2,
                      JAGS.PATH=path.to.JAGS,
                      INPUT.DIR = "<level-2 FILE LOCATION>")

## End(Not run)
```

`calc_fup_uc_point` *Calculate Point Estimates of Fraction Unbound in Plasma (Fup) with Ultracentrifugation (UC) Data (Level-3)*

Description

This function calculates the point estimates for the fraction unbound in plasma (Fup) using mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of chemical Fup using ultracentrifugation (Redgrave et al. 1975). See the Details section for the equation(s) used in the point estimate.

Usage

```
calc_fup_uc_point(
```

```
  FILENAME,
  data.in,
  good.col = "Verified",
  output.res = FALSE,
  sig.figs = 3,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the input level-2 file, "<FILENAME>-fup-UC-Level2.tsv" (if importing from a .tsv file), and/or used to identify the output level-3 file, "<FILENAME>-fup-UC-Level3.tsv" (if exporting).
data.in	(Data Frame) A level-2 data frame generated from the format_fup_uc function with a verification column added by sample_verification. Complement with manual verification if needed.
good.col	(Character) Column name indicating which rows have been verified, data rows valid for analysis are indicated with a "Y". (Defaults to "Verified".)
output.res	(Logical) When set to TRUE, the result table (level-3) will be exported to the user's per-session temporary directory or OUTPUT.DIR (if specified) as a .tsv file. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported result table (level-3). (Note: console print statements are also rounded to specified significant figures.) (Defaults to 3.)
INPUT.DIR	(Character) Path to the directory where the input level-2 file exists. If NULL, looking for the input level-2 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The input to this function should be "level-2" data. Level-2 data is level-1, data formatted with the format_fup_uc function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for analysis.

The should be annotated according to of these types:

Calibration Curve	CC
Ultracentrifugation Aqueous Fraction	AF
Whole Plasma T1h Sample	T1
Whole Plasma T5h Sample	T5

f_{up} is calculated from MS responses as:

$f_{up} = \frac{\sum_{i=1}^{n_A}(r_A * c_{DF})/n_A}{\sum_{i=1}^{n_{T5}}(r_{T5} * c_{DF})/n_{T5}}$

where r_A is Aqueous Fraction Response, c_{DF} is the corresponding Dilution Factor, r_{T5} is T5 Response, n_A is the number of Aqueous Fraction Responses, and n_{T5} is the number of T5 Responses.

If the output level-3 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-3 data frame with one row per chemical, contains chemical identifiers such as preferred compound name, compound name used by the laboratory, EPA's DSSTox Structure ID, calibration, and point estimates for the fraction unbound in plasma (Fup) for all chemicals in the input data frame.

Author(s)

John Wambaugh

References

Redgrave TG, Roberts DCK, West CE (1975). "Separation of plasma lipoproteins by density-gradient ultracentrifugation." *Analytical Biochemistry*, **65**(1–2), 42–49.

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::fup_uc_L2

## scenario 1:
## input level-2 data from the R session and do not export the result table
level3 <- calc_fup_uc_point(data.in = level2, output.res = FALSE)

## scenario 2:
## import level-2 data from a 'tsv' file and export the result table
## Not run:
## Refer to sample_verification help file for how to export level-2 data to a directory.
## Unless a different path is specified in OUTPUT.DIR,
## the result table will be saved to the directory specified in INPUT.DIR.
## Will need to replace FILENAME and INPUT.DIR with name prefix and location of level-2 'tsv'.
level3 <- calc_fup_uc_point(# e.g. replace with "Examples" from "Examples-fup-UC-Level2.tsv"
                            FILENAME="<level-2 FILENAME prefix>",
                            INPUT.DIR = "<level-2 FILE LOCATION>",
                            output.res = TRUE)
```

```
## End(Not run)

## scenario 3:
## import level-2 data from the R session and export the result table to the
## user's temporary directory
## Will need to replace FILENAME with desired level-2 filename prefix.
## Not run:
level3 <- calc_fup_uc_point(# e.g. replace with "MYDATA",
                            FILENAME = "<desired level-2 FILENAME prefix>",
                            data.in = level2,
                            output.res = TRUE)
# To delete, use the following code. For more details, see the link in the
file.remove(list.files(tempdir(), full.names = TRUE,
pattern = "<desired level-2 FILENAME prefix>-fup-UC-Level3.tsv"))

## End(Not run)
```

`check_catalog` *Function to Check Level 0 Data Catalog*

Description

This function is meant to check whether the catalog file is in the anticipated format with required information.

Usage

```
check_catalog(catalog, verbose = TRUE)
```

Arguments

`catalog`	The catalog to be checked, format 'data.frame'.
`verbose`	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Value

(No value returned) Text output indicating whether the level-0 data catalog meets all the necessary requirements in order to auto-extract data from the various source files, or output indicating necessary updates to the data catalog. (NOTE: Nothing is returned if `verbose` is set to FALSE.)

Examples

```
check_catalog(catalog = data.guide) # note the data.guide is not currently in `invitroTKstats`
```

clint_cheminfo *Clint Chemical Information Example Data set*

Description

The chemical ID mapping information from mass spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set contains 7 unique compounds/chemicals.

Usage

```
clint_cheminfo
```

Format

A chemical info data.frame with 7 rows and 6 variables:

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Analyte Name` Name of the test analyte/compound and the name used by the laboratory

`Internal Standard` Name of the internal standard (ISTD)

`Mix` Mix used for the sample

`Compound` Name of the test analyte/compound

`Chem.Lab.ID` Compound as described in the chemistry laboratory

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

clint_L0 *Clint Level-0 Example Data set*

Description

Mass Spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
clint_L0
```

Format

A level-0 data.frame with 247 rows and 16 variables:

`Compound` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.ID` Compound as described in the laboratory

`Date` Date the sample was added to the MS analyzer

`Sample` Sample description used in the laboratory

`Type` Type of Clint sample

`Compound.Conc` Expected (or nominal) concentration of analyte (for calibration curve)

`Peak.Area` Peak area of analyte (target compound)

`ISTD.Peak.Area` Peak area of internal standard (ISTD) compound (pixels)

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`Analysis.Params` Column contains the retention time

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Sample.Text` Additional notes on the sample

`Time` Time when the sample was measured - in hours (h)

`Dilution.Factor` Number of times the sample was diluted

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

clint_L1 *Clint Level-1 Example Data set*

Description

Mass Spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
clint_L1
```

Format

A level-1 data.frame with 229 rows and 24 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of Clint sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to well (uM)

`Hep.Density` The density (units of millions of hepatocytes per mL) hepatocytes in the *in vitro* incubation

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`clint_L2` *Clint Level-2 Example Data set*

Description

Mass Spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
clint_L2
```

Format

A level-2 data.frame with 229 rows and 25 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of Clint sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to well (uM)

`Hep.Density` The density (units of millions of hepatocytes per mL) hepatocytes in the *in vitro* incubation

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

---

clint_L2_heldout *Clint Level-2 Heldout Example Data set*

Description

The unverified level-2 samples from mass spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 2 test analytes/compounds.

Usage

```
clint_L2_heldout
```

Format

A level-2 data.frame with 10 rows and 25 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of Clint sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to well (uM)

`Hep.Density` The density (units of millions of hepatocytes per mL) hepatocytes in the *in vitro* incubation

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

clint_L3 *Clint Level-3 Example Data set*

Description

Mass Spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
clint_L3
```

Format

A level-3 data.frame with 3 rows and 13 variables:

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Calibration` Identifier for mass spectrometry calibration – usually the date

`Clint` Intrinsic hepatic clearance

`Clint.pValue` p-value of estimated `Clint` value

`Fit` Test nominal concentrations

`AIC` Akaike Information Criterion of the linear regression fit

`AIC.Null` Akaike Information Criterion of the exponential decay assuming a constant rate of decay

`Clint.1` Intrinsic hepatic clearance at 1 uM

`Clint.10` Intrinsinc hepatic clearance at 10 uM

`AIC.Sat` Akaike Information Criterion of the exponential decay with a saturation probability

`Sat.pValue` p-value of exponential decay with a saturation probability

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

clint_L4 *Clint Level-4 Example Data set*

Description

Mass Spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
clint_L4
```

Format

A level-4 data.frame with 3 rows and 12 variables:

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Clint.1.Med` Median intrinsic hepatic clearance at 1 uM

`Clint.1.Low` 2.5th quantile of intrinsic hepatic clearance at 1 uM

`Clint.1.High` 97.5th quantile of intrinsic hepatic clearance at 1 uM

`Clint.10.Med` Median of intrinsic hepatic clearance at 10 uM

`Clint.10.Low` 2.5th quantile of intrinsic hepatic clearance at 10 uM

`Clint.10.High` 97.5th quantile of intrinsic hepatic clearance at 1 uM

`Clint.pValue` Probability that a decrease is observed

`Sat.pValue` Saturation probability that a lower `Clint` is observed at a higher concentration

`degrades.pValue` Probability of abiotic degradation

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

clint_PREJAGS *Clint Level-4 PREJAGS arguments*

Description

The arguments given to JAGS for the tested compound during level-4 processing of mass spectrometry measurements of intrinsic hepatic clearance (Clint) for cryopreserved pooled human hepatocytes. Chemicals were per- and poly-fluorinated alkyl substance (PFAS) samples. The experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This list is overwritten for each tested compound. Therefore, only contains arguments given to JAGS for the last tested compound.

Usage

```
clint_PREJAGS
```

Format

A named list with 26 elements:

`obs` `Response` of the "Cvst" sample types for the tested compound

`Test.Nominal.Conc` Unique `Test.Nominal.Conc` values (expected initial concentration) of "Cvst" sample types

`Num.cal` Unique number of `Calibration` values

`Num.obs` Number of `Response` of the "Cvst" sample types for the tested compound

`obs.conc` Indices of the `Test.Nominal.Conc` values that corresponds to the "Cvst" sample types' `Test.Nominal.Conc`

`obs.time` `Time` of the "Cvst" sample types for the tested compound

`obs.cal` Indices of the unique "Cvst" `Calibration` values that corresponds to the "Cvst" sample types' `Calibration`

`obs.Dilution.Factor` `Dilution Factor` of the "Cvst" sample types for the tested compound (number of times the sample was diluted)

`Num.blank.obs` Number of "Blank" sample types for the tested compound

`Blank.obs` `Response` of the "Blank" sample types for the tested compound

`Blank.cal` Indices of the unique "Blank" `Calibration` values that corresponds to the "Blank" sample types' `Calibration`

`Blank.Dilution.Factor` `Dilution Factor` of the "Blank" sample types for the tested compound (number of times the sample was diluted)

`Num.cc` Number of "CC" sample types with non-NA `Test.Compound.Conc` values for the tested compound

`cc.obs.conc` `Test.Compound.Conc` (non-NA) of the "CC" sample types for the tested compound

`cc.obs` `Response` of the "CC" sample types with non-NA `Test.Compound.Conc` for the tested compound

cc.obs.cal Indices of the unique "CC" Calibration values that corresponds to the "CC" sample types' Calibration

cc.obs.Dilution.Factor Dilution Factor of the "CC" sample types (number of times the sample was diluted) with non-NA Test.Compound.Conc for the tested compound

Num.abio.obs Number of "Inactive" samples types for the tested compound

abio.obs Response of the "Inactive" sample types for the tested compound

abio.obs.conc Indices of the Test.Nominal.Conc values that corresponds to the "Inactive" sample types' Test.Nominal.Conc

abio.obs.time Time of the "Inactive" sample types for the tested compound

abio.obs.cal Indices of the unique "Inactive" Calibration values that corresponds to the "Inactive" sample types' Calibration

abio.obs.Dilution.Factor Dilution Factor of the "Inactive" sample types for the tested compound (number of times the sample was diluted)

DECREASE.PROB Prior probability that a chemical will decrease in the assay. (Defaults to 0.5.)

SATURATE.PROB Prior probability that a chemicals rate of metabolism will decrease between 1 and 10 uM. (Defaults to 0.25.)

DEGRADE.PROB Prior probability that a chemical will be unstable (degrade abiotically) in the assay. (Defaults to 0.05.)

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

---

create_catalog *Function to create a catalog of level 0 files to be merged.*

Description

This function is meant for creating a catalog of all level 0 data files listed that will be merged with the 'merge_level0' function. All arguments are required, with exception of 'additional.info'.

Usage

```
create_catalog(
  file,
  sheet,
  skip.rows,
  date,
  compound,
```

```
  istd,
  col.names.loc,
  sample,
  type,
  peak,
  istd.peak,
  conc,
  analysis.param,
  num.rows = NULL,
  additional.info = NULL,
  verbose = TRUE
)
```

Arguments

file	(*character vector*) Vector of character strings with the file names of level 0 data.
sheet	(*character vector*) Vector of character strings containing the sheet name with MS data.
skip.rows	(*numeric vector*) Numeric vector containing the number of rows to skip in data file.
date	(*character vector*) Vector of character strings containing the date of data collection, format "MMDDYY". "MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year.
compound	(*character vector*) Vector of character strings with the relevant chemical identifier.
istd	(*character vector*) Vector of character strings with the internal standard.
col.names.loc	(*numeric vector*) Numeric vector containing the row locations of the column names.
sample	(*character vector*) Vector of character strings with column names containing samples.
type	(*character vector*) Vector of character strings with column names containing type information.
peak	(*character vector*) Vector of character strings with the column names containing mass spectrometry (MS) peak data.
istd.peak	(*character vector*) Vector of character strings with column names containing internal standard (ITSD) peak data.
conc	(*character vector*) Vector of character strings with column names containing exposure concentration data.
analysis.param	(*character vector*) Vector of character strings with column names containing analysis parameters.
num.rows	(*numeric vector*) Numeric vector containing the number of rows with data to be pulled. (Default is NULL.)
additional.info	(*list* or *data.frame*) Named list or data.frame of additional columns to include in the catalog. Additional columns should follow the nomenclature of "<Fill-in>.ColName" if indicating column names with information to pull, otherwise a

short name. All spaces in additional column names should be designated with a period, "." . (Default is NULL, i.e. no additional columns.)

verbose (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Value

(*data.frame*) A catalog containing information about the source level-0 data file to enable proper 'auto-extraction' of data. Additionally, the catalog contains other relevant meta-data fields describing when, how, what, etc. of the assay that collected the level-0 data.

See Also

merge_level0

Examples

```
create_catalog(
  file = "testME.xlsx",sheet = "3",skip.rows = 0,
  date = "112723",compound = "80-05-7",
  istd = "Chemical A", col.names.loc = 1,
  sample = "Sample.Name",type = "Type",
  peak = "Response.Area",istd.peak = "ISTD.Peak.Area",
  conc = "Intended.Concentration",analysis.param = "A,B,C"
)
```

---

create_chem_table *Creates a Standardized Data Table of Chemical Identities*

Description

This function creates a data frame summarizing chemical identifiers used for each tested chemical in MS data. Each row in the resulting data frame provides EPA's DSSTox Structure ID (DTXSID), preferred compound name, and the name used by the laboratory.

Usage

```
create_chem_table(
  input.table,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  lab.compound.col = "Lab.Compound.Name",
  verbose = TRUE
)
```

Arguments

input.table	(Data Frame) A data frame containing mass-spectrometry peak areas, indication of chemical identity, and analytical chemistry methods. It should contain columns with names specified by the following arguments:
dtxsid.col	(Character) Column name of input.table containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)
compound.col	(Character) Column name of input.table containing the test compound. (Defaults to "Compound.Name".)
lab.compound.col	(Character) Column name of input.table containing the test compound name used by the laboratory. (Defaults to "Lab.Compound.Name".)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Value

A data frame containing the chemical identifiers for all unique chemicals in the input data frame. Each row maps a unique chemical, indicated by the DTXSID, to all the preferred compound names and all chemical names used by the laboratory referenced in the input data frame.

Author(s)

John Wambaugh

Examples

```
library(invitroTKstats)
# Smeltz et al. (2020) data:
##  Clint ##
create_chem_table(
  input.table = invitroTKstats::clint_cheminfo,
  dtxsid.col = "DTXSID",
  compound.col = "Compound",
  lab.compound.col = "Chem.Lab.ID"
  )
## Fup RED ##
create_chem_table(
  input.table = invitroTKstats::fup_red_cheminfo,
  dtxsid.col = "DTXSID",
  compound.col = "Compound",
  lab.compound.col = "Chem.Lab.ID"
  )
## Fup UC ##
create_chem_table(
  input.table = invitroTKstats::fup_uc_cheminfo,
  dtxsid.col = "DTXSID",
  compound.col = "Compound",
  lab.compound.col = "Chem.Lab.ID"
  )
# Honda et al. () data:
## Caco2 ##
```

```
create_chem_table(
  input.table = invitroTKstats::caco2_cheminfo,
  dtxsid.col = "DTXSID",
  compound.col = "PREFERRED_NAME",
  lab.compound.col = "test_article"
  )
```

`create_method_table` *Creates a Standardized Data Table for Chemical Analysis Methods*

Description

This function extracts the chemical analysis methods from a set of MS data and returns a data frame with each row representing a unique chemical-method pair. (Unique chemical identified by DTXSID.) Each row contains all compound names, analysis parameters, analysis instruments, and internal standards used for each chemical-method pair.

Usage

```
create_method_table(
  input.table,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  istd.name.col = "ISTD.Name",
  analysis.method.col = "Analysis.Method",
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters.col = "Analysis.Parameters",
  verbose = TRUE
)
```

Arguments

`input.table`	(Data Frame) A level-1 or level-2 data frame containing mass-spectrometry peak areas, indication of chemical identity, and analytical chemistry methods. It should contain columns with names specified by the following arguments:
`dtxsid.col`	(Character) Column name of `input.table` containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)
`compound.col`	(Character) Column name of `input.table` containing the test compound. (Defaults to "Compound.Name".)
`istd.name.col`	(Character) Column name of `input.table` containing identity of the internal standard. (Defaults to "ISTD.Name".)
`analysis.method.col`	(Character) Column name of `input.table` containing the analytical chemistry analysis method, typically "LCMS" or "GCMS", liquid or gas chromatography mass spectrometry, respectively. (Defaults to "Analysis.Method".)

`analysis.instrument.col`
: (Character) Column name of `input.table` containing the instrument used for chemical analysis. For example, "Agilent 6890 GC with model 5973 MS". (Defaults to "Analysis.Instrument".)

`analysis.parameters.col`
: (Character) Column name of `input.table` containing the parameters used to identify the compound on the chemical analysis instrument. For example, "Negative Mode, 221.6/161.6, -DPb=26, FPc=-200, EPd=-10, CEe=-20, CXPf=-25.0". (Defaults to "Analysis.Parameters".)

`verbose`
: (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Value

A data frame with one row per chemical-method pair containing information on analysis parameters, instruments, internal standards, and compound identifiers used for each pair.

Author(s)

John Wambaugh

Examples

```
library(invitroTKstats)
# Smeltz et al. (2020) data:
##  Clint ##
create_method_table(
  input.table = invitroTKstats::clint_L1,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  istd.name.col = "ISTD.Name",
  analysis.method.col = "Analysis.Method",
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters.col = "Analysis.Parameters"
  )
## Fup RED ##
create_method_table(
  input.table = invitroTKstats::fup_red_L1,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  istd.name.col = "ISTD.Name",
  analysis.method.col = "Analysis.Method",
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters.col = "Analysis.Parameters"
  )
## Fup UC ##
create_method_table(
  input.table = invitroTKstats::fup_uc_L1,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  istd.name.col = "ISTD.Name",
  analysis.method.col = "Analysis.Method",
```

```
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters.col = "Analysis.Parameters"
  )
# Honda et al. () data:
## Caco2 ##
create_method_table(
  input.table = invitroTKstats::caco2_L1,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  istd.name.col = "ISTD.Name",
  analysis.method.col = "Analysis.Method",
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters.col = "Analysis.Parameters"
  )
```

`extract_level1_fup_uc` *Extract level 1 ultracentrifugation (Redgrave et al. 1975) data from wide level 0 file*

Description

This function extracts data from a Microsoft Excel file containing many columns corresponding to different types of data.

Usage

```
extract_level1_fup_uc(
  data.set,
  chem.name,
  area.col.num,
  ISTD.name,
  ISTD.offset = 2,
  analysis.method = "GC",
  instrument = "Something or Other 3000",
  inst.param.offset = -3,
  conc.offset = -2,
  area.base = "Area...",
  inst.param.base = "RT...",
  conc.base = "Final Conc....",
  id.cols = c("Name", "Data File", "Acq. Date-Time"),
  type.indicator.col = "Name",
  AF.type.str = "AF",
  T1.type.str = "T1",
  T5.type.str = "T5",
  CC.type.str = "CC"
)
```

Arguments

data.set	(Data Frame) A data frame containing a sheet of data for conversion.
chem.name	(Character) A string giving the lab name of the chemical analyzed. The value provided is used for all rows in the output data frame.
area.col.num	(Numeric) An integer indicating which column of data.set contains the MS feature area for the chemical.
ISTD.name	(Character) A string indicating the internal standard used. The value provided is used for all rows in the output data frame.
ISTD.offset	(Numeric) An integer indicating how many columns difference there is between the chemical of study MS area and the ISTD MS area. (Defaults to 2.)
analysis.method	(Character) A string describing the chemical analysis method. The value provided is used for all rows in the output data frame. (Defaults to "GC", that is gas chromatography.)
instrument	(Character) A string describing the instrument used for chemical analysis. The value provided is used for all rows in the output data frame. (Defaults to "Something or Other 3000".)
inst.param.offset	(Numeric) An integer indicating the difference in the number of columns between the MS peak area and the column giving the instrument parameters. (Defaults to -3.)
conc.offset	(Numeric) An integer indicating the difference in the number of columns between the MS peak area and the column giving the intended concentration for calibration curves. (Defaults to -2.)
area.base	(Character) A character string used for forming the name of MS feature area column names (used for both test chemical and ISTD). (Defaults to "Area...".)
inst.param.base	(Character) A character string used for forming the name of the chemical analysis instrument parameter column name. (Defaults to "RT...".)
conc.base	(Character) A character string used for forming the name of the calibration curve intended concentration column name. (Defaults to "Final Conc....".)
id.cols	(Character Vector) A vector of character strings used for identifying each sample. (Defaults to c("Name", "Data File", "Type", "Acq. Date-Time").)
type.indicator.col	(Character) A character string indicating which column of data.set contains the type of observation. (Defaults to "Name".)
AF.type.str	(Character) String used to annotate observation of this type: Aqueous Fraction. (Defaults to "AF".)
T1.type.str	(Character) String used to annotate observation of this type: Whole Plasma T1h Sample. (Defaults to "T1".)
T5.type.str	(Character) String used to annotate observation of this type: Whole Plasma T5h Sample. (Defaults to "T5".)
CC.type.str	(Character) String used to annotate observation of this type: Calibration Curve. (Defaults to "CC".)

Details

The data frame of observations should be annotated according to of these types:

Calibration Curve	CC
Ultracentrifugation Aqueous Fraction	AF
Whole Plasma T1h Sample	T1
Whole Plasma T5h Sample	T5

Value

`data.frame` A data.frame in standardized "level1" format

Author(s)

John Wambaugh

References

Redgrave TG, Roberts DCK, West CE (1975). "Separation of plasma lipoproteins by density-gradient ultracentrifugation." *Analytical Biochemistry*, **65**(1–2), 42–49.

`format_caco2` *Creates a Standardized Data Frame with Caco-2 Data (Level-1)*

Description

This function formats data describing mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of membrane permeability using Caco-2 cells (Hubatsch et al. 2007). The input data frame is organized into a standard set of columns and is written to a tab-separated text file.

Usage

```
format_caco2(
  FILENAME = "MYDATA",
  data.in,
  sample.col = "Lab.Sample.Name",
  lab.compound.col = "Lab.Compound.Name",
  dtxsid.col = "DTXSID",
  date = NULL,
  date.col = "Date",
  compound.col = "Compound.Name",
  area.col = "Area",
  istd.col = "ISTD.Area",
  type.col = "Type",
  direction.col = "Direction",
```

```
  membrane.area = NULL,
  membrane.area.col = "Membrane.Area",
  receiver.vol.col = "Vol.Receiver",
  donor.vol.col = "Vol.Donor",
  test.conc = NULL,
  test.conc.col = "Test.Compound.Conc",
  cal = NULL,
  cal.col = "Cal",
  dilution = NULL,
  dilution.col = "Dilution.Factor",
  time = NULL,
  time.col = "Time",
  istd.name = NULL,
  istd.name.col = "ISTD.Name",
  istd.conc = NULL,
  istd.conc.col = "ISTD.Conc",
  test.nominal.conc = NULL,
  test.nominal.conc.col = "Test.Target.Conc",
  biological.replicates = NULL,
  biological.replicates.col = "Biological.Replicates",
  technical.replicates = NULL,
  technical.replicates.col = "Technical.Replicates",
  analysis.method = NULL,
  analysis.method.col = "Analysis.Method",
  analysis.instrument = NULL,
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters = NULL,
  analysis.parameters.col = "Analysis.Parameters",
  note.col = "Note",
  level0.file = NULL,
  level0.file.col = "Level0.File",
  level0.sheet = NULL,
  level0.sheet.col = "Level0.Sheet",
  output.res = FALSE,
  save.bad.types = FALSE,
  sig.figs = 5,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the output level-1 file. "<FILENAME>-Caco-2-Level1.tsv", and/or used to identify the input level-0 file, "<FILENAME>-Caco-2-Level0.tsv" if importing from a .tsv file. (Defaults to "MYDATA".)
data.in	(Data Frame) A level-0 data frame containing mass-spectrometry peak areas, indication of chemical identity, and measurement type. The data frame should

contain columns with names specified by the following arguments:

`sample.col` (Character) Column name of `data.in` containing the unique mass spectrometry (MS) sample name used by the laboratory. (Defaults to "Lab.Sample.Name".)

`lab.compound.col` (Character) Column name of `data.in` containing the test compound name used by the laboratory. (Defaults to "Lab.Compound.Name".)

`dtxsid.col` (Character) Column name of `data.in` containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)

`date` (Character) The laboratory measurement date, format "MMDDYY" where "MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year. (Defaults to `NULL`.) (Note: Single entry only, use only if all data were collected on the same date.)

`date.col` (Character) Column name containing `date` information. (Defaults to "Date".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `date`.)

`compound.col` (Character) Column name of `data.in` containing the test compound. (Defaults to "Compound.Name".)

`area.col` (Character) Column name of `data.in` containing the target analyte (that is, the test compound) MS peak area. (Defaults to "Area".)

`istd.col` (Character) Column name of `data.in` containing the MS peak area for the internal standard. (Defaults to "ISTD.Area".)

`type.col` (Character) Column name of `data.in` containing the sample type (see table under Details). (Defaults to "Type".)

`direction.col` (Character) Column name of `data.in` containing the direction of the Caco-2 permeability experiment: either apical donor to basolateral receiver (AtoB), or basolateral donor to apical receiver (BtoA). (Defaults to "Direction".)

`membrane.area` (Numeric) The area of the Caco-2 monolayer (in cm^2). (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds have the same area for the Caco-2 monolayer.)

`membrane.area.col` (Character) Column name containing `membrane.area` information. (Defaults to "Membrane.Area".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `membrane.area`.)

`receiver.vol.col` (Character) Column name of `data.in` containing the media volume (in cm^3) of the receiver portion of the Caco-2 experimental well. (Defaults to "Vol.Receiver".)

`donor.vol.col` (Character) Column name of `data.in` containing the media volume (in cm^3) of the donor portion of the Caco-2 experimental well where the test chemical is added. (Defaults to "Vol.Donor".)

`test.conc` (Numeric) The standard test chemical concentration for the Caco-2 assay. (Defaults to `NULL`.) (Note: Single entry only, use only if the same standard concentration was used for all tested compounds.)

`test.conc.col` (Character) Column name containing `test.conc` information. (Defaults to "Test.Compound.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `test.conc`.)

`cal` (Character) MS calibration the samples were based on. Typically, this uses indices or dates to represent if the analyses were done on different machines on the same day or on different days with the same MS analyzer. (Defaults to `NULL`.) (Note: Single entry only, use only if all data were collected based on the same calibration.)

`cal.col` (Character) Column name containing `cal` information. (Defaults to "Cal".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `cal`.)

`dilution` (Numeric) Number of times the sample was diluted before MS analysis. (Defaults to `NULL`.) (Note: Single entry only, use only if all samples underwent the same number of dilutions.)

`dilution.col` (Character) Column name containing `dilution` information. (Defaults to "Dilution.Factor".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `dilution`.)

`time` (Numeric) The amount of time (in hours) before the receiver and donor compartments are measured. (Defaults to `NULL`.)

`time.col` (Character) Column name containing `meas.time` information. (Defaults to "Time".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `meas.time`.)

`istd.name` (Character) The identity of the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds use the same internal standard.)

`istd.name.col` (Character) Column name containing `istd.name` information. (Defaults to "ISTD.Name".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `istd.name`.)

`istd.conc` (Numeric) The concentration for the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds have the same internal standard concentration.)

`istd.conc.col` (Character) Column name containing `istd.conc` information. (Defaults to "ISTD.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `istd.conc`.)

`test.nominal.conc`
(Numeric) The nominal concentration added to the donor compartment at time 0. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds used the same concentration at time 0.

`test.nominal.conc.col`
(Character) Column name containing `test.nominal.conc` information. (Defaults to "Test.Target.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `test.nominal.conc`.)

`biological.replicates`
(Character) Replicates with the same analyte. Typically, this uses numbers or letters to index. (Defaults to `NULL`.) (Note: Single entry only, use only if none of the test compounds have replicates.)

`biological.replicates.col`
(Character) Column name of `data.in` containing the number or the indices of replicates with the same analyte. (Defaults to "Biological.Replicates".) (Note:

data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in biological.replicates.)

technical.replicates
: (Character) Repeated measurements from one sample. Typically, this uses numbers or letters to index. (Defaults to NULL.) (Note: Single entry only, use only if none of the test compounds have replicates.)

technical.replicates.col
: (Character) Column name of data.in containing the number or the indices of replicates taken from the one sample. (Defaults to "Technical.Replicates".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in technical.replicates.)

analysis.method
: (Character) The analytical chemistry analysis method, typically "LCMS" or "GCMS", liquid chromatography or gas chromatography–mass spectrometry, respectively. (Defaults to NULL.) (Note: Single entry only, use only if the same method was used for all tested compounds.)

analysis.method.col
: (Character) Column name containing analysis.method information. (Defaults to "Analysis.Method".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.method.)

analysis.instrument
: (Character) The instrument used for chemical analysis, for example "Agilent 6890 GC with model 5973 MS". (Defaults to NULL.) (Note: Single entry only, use only if the same instrument was used for all tested compounds.)

analysis.instrument.col
: (Character) Column name containing analysis.instrument information. (Defaults to "Analysis.Instrument".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.instrument.)

analysis.parameters
: (Character) The parameters used to identify the compound on the chemical analysis instrument, for example "Negative Mode, 221.6/161.6, -DPb=26, FPc=-200, EPd=-10, CEe=-20, CXPf=-25.0". (Defaults to NULL.) (Note: Single entry only, use only if the same parameters were used for all tested compounds.)

analysis.parameters.col
: (Character) Column name containing analysis.parameters information. (Defaults to "Analysis.Parameters".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.parameters.)

note.col
: (Character) Column name of data.in containing additional notes on test compounds. (Defaults to "Note").

level0.file
: (Character) The level-0 file from which the data.in were obtained. (Defaults to NULL.) (Note: Single entry only, use only if all rows in data.in were obtained from the same level-0 file.)

level0.file.col
: (Character) Column name containing level0.file information. (Defaults to

"Level0.File".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.file`.)

`level0.sheet` (Character) The specific sheet name of level-0 file from which the `data.in` is obtained from, if the level-0 file is an Excel workbook. (Defaults to `NULL`.) (Note: Single entry only, use only if all rows in `data.in` were obtained from the same sheet in the same level-0 file.)

`level0.sheet.col` (Character) Column name containing `level0.sheet` information. (Defaults to "Level0.Sheet".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.sheet`.)

`output.res` (Logical) When set to `TRUE`, the result table (level-1) will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)

`save.bad.types` (Logical) When set to `TRUE`, export data removed due to inappropriate sample types. See the Detail section for the required sample types. (Defaults to `FALSE`.)

`sig.figs` (Numeric) The number of significant figures to round the exported result table (level-1). (Defaults to 5.)

`INPUT.DIR` (Character) Path to the directory where the input level-0 file exists. If `NULL`, looking for the input level-0 file in the current working directory. (Defaults to `NULL`.)

`OUTPUT.DIR` (Character) Path to the directory to save the output file. If `NULL`, the output file will be saved to the user's per-session temporary directory or `INPUT.DIR` if specified. (Defaults to `NULL`.)

`verbose` (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

In this experiment an *in vitro* well is separated into two by a membrane composed of a monolayer of Caco-2 cells. A test chemical is added to either the apical or basolateral side of of the monolayer at time 0, and after a set time samples are taken from both the "donor" (side where the test chemical was added) and the "receiver" side. Depending on the direction of the test the donor side can be either apical or basolateral.

The data frame of observations should be annotated according to direction (either apical to basolateral – "AtoB" – or basolateral to apical – "BtoA") and type of concentration measured:

Blank with no chemical added	Blank
Target concentration added to donor compartment at time 0 (C0)	D0
Donor compartment at end of experiment	D2
Receiver compartment at end of experiment	R2

Chemical concentration is calculated qualitatively as a response and returned as a column in the output data frame:

Response <- AREA / ISTD.AREA * ISTD.CONC

If the output level-1 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-1 data frame with a standardized format containing a standardized set of columns and column names with membrane permeability data from a Caco-2 assay.

Author(s)

John Wambaugh

References

Hubatsch I, Ragnarsson EG, Artursson P (2007). "Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers." *Nature protocols*, **2**(9), 2111–2119.

Examples

```
## Load example level-0 data and do not export the result table
level0 <- invitroTKstats::caco2_L0
level1 <- format_caco2(data.in = level0,
                       sample.col = "Sample",
                       lab.compound.col = "Lab.Compound.ID",
                       compound.col = "Compound",
                       area.col = "Peak.Area",
                       istd.col = "ISTD.Peak.Area",
                       membrane.area = 0.11,
                       test.conc.col = "Compound.Conc",
                       cal = 1,
                       time = 2,
                       istd.conc = 1,
                       test.nominal.conc = 10,
                       biological.replicates = 1,
                       technical.replicates = 1,
                       analysis.method.col = "Analysis.Params",
                       analysis.instrument = "Agilent.GCMS",
                       analysis.parameters = "Unknown",
                       note.col = NULL,
                       output.res = FALSE
)
```

format_clint *Creates a Standardized Data Frame with Hepatocyte Clearance Data (Level-1)*

Description

This function formats data describing mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of chemical stability when incubated with suspended hepatocytes (Shibata et al. 2002). Disappearance of the chemical over time is assumed to be due to metabolism by the hepatocytes. The input data frame is organized into a standard set of columns and is written to a tab-separated text file.

Usage

```
format_clint(
  FILENAME = "MYDATA",
  data.in,
  sample.col = "Lab.Sample.Name",
  date = NULL,
  date.col = "Date",
  compound.col = "Compound.Name",
  dtxsid.col = "DTXSID",
  lab.compound.col = "Lab.Compound.Name",
  type.col = "Sample.Type",
  density = NULL,
  density.col = "Hep.Density",
  cal = NULL,
  cal.col = "Cal",
  dilution = NULL,
  dilution.col = "Dilution.Factor",
  time = NULL,
  time.col = "Time",
  istd.col = "ISTD.Area",
  istd.name = NULL,
  istd.name.col = "ISTD.Name",
  istd.conc = NULL,
  istd.conc.col = "ISTD.Conc",
  test.conc = NULL,
  test.conc.col = "Test.Compound.Conc",
  test.nominal.conc = NULL,
  test.nominal.conc.col = "Test.Target.Conc",
  area.col = "Area",
  biological.replicates = NULL,
  biological.replicates.col = "Biological.Replicates",
  technical.replicates = NULL,
  technical.replicates.col = "Technical.Replicates",
  analysis.method = NULL,
```

```
  analysis.method.col = "Analysis.Method",
  analysis.instrument = NULL,
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters = NULL,
  analysis.parameters.col = "Analysis.Parameters",
  note.col = "Note",
  level0.file = NULL,
  level0.file.col = "Level0.File",
  level0.sheet = NULL,
  level0.sheet.col = "Level0.Sheet",
  output.res = FALSE,
  save.bad.types = FALSE,
  sig.figs = 5,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the output level-1 file. "<FILENAME>-Clint-Level1.tsv", and/or used to identify the input level-0 file, "<FILENAME>-Clint-Level0.tsv" if importing from a .tsv file. (Defaults to "MYDATA").
data.in	(Data Frame) A level-0 data frame or a matrix containing mass-spectrometry peak areas, indication of chemical identity, and measurement type. The data frame should contain columns with names specified by the following arguments:
sample.col	(Character) Column name of data.in containing the unique mass spectrometry (MS) sample name used by the laboratory. (Defaults to "Lab.Sample.Name".)
date	(Character) The laboratory measurement date, format "MMDDYY" where "MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year. (Defaults to NULL.) (Note: Single entry only, use only if all data were collected on the same date.)
date.col	(Character) Column name containing date information. (Defaults to "Date".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in date.)
compound.col	(Character) Column name of data.in containing the test compound. (Defaults to "Compound.Name".)
dtxsid.col	(Character) Column name of data.in containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)
lab.compound.col	(Character) Column name of data.in containing the test compound name used by the laboratory. (Defaults to "Lab.Compound.Name".)
type.col	(Character) Column name of data.in containing the sample type (see table under Details). (Defaults to "Sample.Type".)
density	(Numeric) The density (units of millions of hepatocytes per mL) hepatocytes in the *in vitro* incubation. (Defaults to NULL.) (Note: Single entry only, use only if all tested compounds have the same density.)

density.col (Character) Column name containing density information. (Defaults to "Hep.Density".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in density.)

cal (Character) MS calibration the samples were based on. Typically, this uses indices or dates to represent if the analyses were done on different machines on the same day or on different days with the same MS analyzer. (Defaults to NULL.) (Note: Single entry only, use only if all data were collected based on the same calibration.)

cal.col (Character) Column name containing cal information. (Defaults to "Cal".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in cal.)

dilution (Numeric) Number of times the sample was diluted before MS analysis. (Defaults to NULL.) (Note: Single entry only, use only if all samples underwent the same number of dilutions.)

dilution.col (Character) Column name containing dilution information. (Defaults to "Dilution.Factor".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in dilution.)

time (Numeric) Time of the measurement (in minutes) since the test chemicals was introduced into the hepatocyte incubation. (Defaults to NULL.) (Note: Single entry only, use only if all measurements were taken after the same amount of time.)

time.col (Character) Column name containing time information. (Defaults to "Time".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in time.)

istd.col (Character) Column name of data.in containing the MS peak area for the internal standard. (Defaults to "ISTD.Area".)

istd.name (Character) The identity of the internal standard. (Defaults to NULL.) (Note: Single entry only, use only if all tested compounds use the same internal standard.)

istd.name.col (Character) Column name containing istd.name information. (Defaults to "ISTD.Name".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in istd.name.)

istd.conc (Numeric) The concentration for the internal standard. (Defaults to NULL.) (Note: Single entry only, use only if all tested compounds have the same internal standard concentration.)

istd.conc.col (Character) Column name containing istd.conc information. (Defaults to "ISTD.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in istd.conc.)

test.conc (Numeric) The standard test chemical concentration for the intrinsic clearance assay. (Defaults to NULL.) (Note: Single entry only, use only if the same standard concentration was used for all tested compounds.)

test.conc.col (Character) Column name containing test.conc information. (Defaults to "Test.Compound.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in test.conc.)

`test.nominal.conc`
: (Numeric) The nominal concentration added to the well at time 0. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds used the same concentration at time 0.)

`test.nominal.conc.col`
: (Character) Column name containing `test.nominal.conc` information. (Defaults to "Test.Target.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `test.nominal.conc`.)

`area.col`
: (Character) Column name of `data.in` containing the target analyte (that is, the test compound) MS peak area. (Defaults to "Area".)

`biological.replicates`
: (Character) Replicates with the same analyte. Typically, this uses numbers or letters to index. (Defaults to `NULL`.) (Note: Single entry only, use only if none of the test compounds have replicates.)

`biological.replicates.col`
: (Character) Column name of `data.in` containing the number or the indices of replicates with the same analyte. (Defaults to "Biological.Replicates".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `biological.replicates`.)

`technical.replicates`
: (Character) Repeated measurements from one sample. Typically, this uses numbers or letters to index. (Defaults to `NULL`.) (Note: Single entry only, use only if none of the test compounds have replicates.)

`technical.replicates.col`
: (Character) Column name of `data.in` containing the number or the indices of replicates taken from the one sample. (Defaults to "Technical.Replicates".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `technical.replicates`.)

`analysis.method`
: (Character) The analytical chemistry analysis method, typically "LCMS" or "GCMS", liquid chromatography or gas chromatography–mass spectrometry, respectively. (Defaults to `NULL`.) (Note: Single entry only, use only if the same method was used for all tested compounds.)

`analysis.method.col`
: (Character) Column name containing `analysis.method` information. (Defaults to "Analysis.Method".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.method`.)

`analysis.instrument`
: (Character) The instrument used for chemical analysis, for example "Waters Xevo TQ-S micro (QEB0036)". (Defaults to `NULL`.) (Note: Single entry only, use only if the same instrument was used for all tested compounds.)

`analysis.instrument.col`
: (Character) Column name containing `analysis.instrument` information. (Defaults to "Analysis.Instrument".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.instrument`.)

`analysis.parameters`
: (Numeric) The parameters used to identify the compound on the chemical analysis instrument. (Defaults to `NULL`.) (Note: Single entry only, use only if the same parameters were used for all tested compounds.)

`analysis.parameters.col`
: (Character) Column name containing `analysis.parameters` information. (Defaults to "Analysis.Parameters".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.parameters`.)

`note.col`
: (Character) Column name of `data.in` containing additional notes on test compounds. (Defaults to "Note").

`level0.file`
: (Character) The level-0 file from which the `data.in` were obtained. (Defaults to `NULL`.) (Note: Single entry only, use only if all rows in data.in were obtained from the same level-0 file.)

`level0.file.col`
: (Character) Column name containing `level0.file` information. (Defaults to "Level0.File".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.file`.)

`level0.sheet`
: (Character) The specific sheet name of level-0 file from which the `data.in` is obtained from, if the level-0 file is an Excel workbook. (Defaults to `NULL`.) (Note: Single entry only, use only if all rows in `data.in` were obtained from the same sheet in the same level-0 file.)

`level0.sheet.col`
: (Character) Column name containing `level0.sheet` information. (Defaults to "Level0.Sheet".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.sheet`.)

`output.res`
: (Logical) When set to `TRUE`, the result table (level-1) will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)

`save.bad.types`
: (Logical) When set to `TRUE`, export data removed due to inappropriate sample types. See the Detail section for the required sample types. (Defaults to `FALSE`.)

`sig.figs`
: (Numeric) The number of significant figures to round the exported result table (level-1). (Defaults to 5.)

`INPUT.DIR`
: (Character) Path to the directory where the input level-0 file exists. If `NULL`, looking for the input level-0 file in the current working directory. (Defaults to `NULL`.)

`OUTPUT.DIR`
: (Character) Path to the directory to save the output file. If `NULL`, the output file will be saved to the user's per-session temporary directory or `INPUT.DIR` if specified. (Defaults to `NULL`.)

`verbose`
: (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The data frame of observations should be annotated according to these types:

Blank	Blank
Hepatocyte incubation concentration	Cvst
Inactivated Hepatocytes	Inactive
Calibration Curve	CC

Chemical concentration is calculated qualitatively as a response and returned as a column in the output data frame:

Response <- AREA / ISTD.AREA * ISTD.CONC

If the output level-1 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

NOTE: For the estimation of Cl~int~ the 'test.conc' and 'test.conc.col' are not used within the calculations currently. However, to maintain consistency with other assays and for the use case that a calibration curve may be part of the estimation in future this was retained. We suggest that if the users do not have a corresponding compound column to set 'test.conc' to 'NA' or use the next most appropriate value/level-0 column name.

Value

A level-1 data frame with a standardized format containing a standardized set of columns and column names with hepatic clearance data for a variety of chemicals.

Author(s)

John Wambaugh

References

Shibata Y, Takahashi H, Chiba M, Ishii Y (2002). "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition*, **30**(8), 892–896.

Examples

```
## Load the example level-0 data
level0 <- invitroTKstats::clint_L0

## Run it through level-1 processing function
## This example shows the use of the data.in argument which allows users to pass
## in a data frame from the R session.
## If the input level-0 data exists in an external file such as a .tsv file,
## users may import it using INPUT.DIR to specify the path and FILENAME
## to specify the file name. See documentation for details.
level1 <- format_clint(data.in = level0,
```

```
sample.col ="Sample",
date.col="Date",
compound.col="Compound",
lab.compound.col="Lab.Compound.ID",
type.col="Type",
dilution.col="Dilution.Factor",
cal=1,
istd.conc = 10/1000,
istd.col= "ISTD.Peak.Area",
area.col = "Peak.Area",
density = 0.5,
test.nominal.conc = 1,
biological.replicates = 1,
test.conc.col="Compound.Conc",
time.col = "Time",
analysis.method = "LCMS",
analysis.instrument = "Unknown",
analysis.parameters.col = "Analysis.Params",
note="Sample Text",
output.res = FALSE
)
```

`format_fup_red` *Creates a Standardized Data Frame with Rapid Equilibrium Dialysis (RED) Plasma Protein Binding (PPB) Data (Level-1)*

Description

This function formats data describing mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of chemical fraction unbound in plasma using rapid equilibrium dialysis (Waters et al. 2008). The input data frame is organized into a standard set of columns and written to a tab-separated text file.

Usage

```
format_fup_red(
  FILENAME = "MYDATA",
  data.in,
  sample.col = "Lab.Sample.Name",
  date = NULL,
  date.col = "Date",
  compound.col = "Compound.Name",
  dtxsid.col = "DTXSID",
  lab.compound.col = "Lab.Compound.Name",
  type.col = "Sample.Type",
  cal = NULL,
  cal.col = "Cal",
  dilution = NULL,
```

```
  dilution.col = "Dilution.Factor",
  time = NULL,
  time.col = "Time",
  istd.col = "ISTD.Area",
  istd.name = NULL,
  istd.name.col = "ISTD.Name",
  istd.conc = NULL,
  istd.conc.col = "ISTD.Conc",
  test.nominal.conc = NULL,
  test.nominal.conc.col = "Test.Target.Conc",
  plasma.percent = NULL,
  plasma.percent.col = "Plasma.Percent",
  test.conc = NULL,
  test.conc.col = "Test.Compound.Conc",
  area.col = "Area",
  biological.replicates = NULL,
  biological.replicates.col = "Biological.Replicates",
  technical.replicates = NULL,
  technical.replicates.col = "Technical.Replicates",
  analysis.method = NULL,
  analysis.method.col = "Analysis.Method",
  analysis.instrument = NULL,
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters = NULL,
  analysis.parameters.col = "Analysis.Parameters",
  note.col = "Note",
  level0.file = NULL,
  level0.file.col = "Level0.File",
  level0.sheet = NULL,
  level0.sheet.col = "Level0.Sheet",
  output.res = FALSE,
  save.bad.types = FALSE,
  sig.figs = 5,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the output level-1 file. "<FILENAME>-fup-RED-Level1.tsv", and/or used to identify the input level-0 file, "<FILENAME>-fup-RED-Level0.tsv" if importing from a .tsv file. (Defaults to "MYDATA".)
data.in	(Data Frame) A level-0 data frame containing mass-spectrometry peak areas, indication of chemical identity, and measurement type. The data frame should contain columns with names specified by the following arguments:
sample.col	(Character) Column name of data.in containing the unique mass spectrometry (MS) sample name used by the laboratory. (Defaults to "Lab.Sample.Name".)

`date` (Character) The laboratory measurement date, format "MMDDYY" where "MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year. (Defaults to `NULL`.) (Note: Single entry only, use only if all data were collected on the same date.)

`date.col` (Character) Column name containing `date` information. (Defaults to "Date".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `date`.)

`compound.col` (Character) Column name of `data.in` containing the test compound. (Defaults to "Compound.Name".)

`dtxsid.col` (Character) Column name of `data.in` containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)

`lab.compound.col` (Character) Column name of `data.in` containing the test compound name used by the laboratory. (Defaults to "Lab.Compound.Name".)

`type.col` (Character) Column name of `data.in` containing the sample type (see table under Details). (Defaults to "Sample.Type".)

`cal` (Character) MS calibration the samples were based on. Typically, this uses indices or dates to represent if the analyses were done on different machines on the same day or on different days with the same MS analyzer. (Defaults to `NULL`.) (Note: Single entry only, use only if all data were collected based on the same calibration.)

`cal.col` (Character) Column name containing `cal` information. (Defaults to "Cal".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `cal`.)

`dilution` (Numeric) Number of times the sample was diluted before MS analysis. (Defaults to `NULL`.) (Note: Single entry only, use only if all samples underwent the same number of dilutions.)

`dilution.col` (Character) Column name containing `dilution` information. (Defaults to "Dilution.Factor".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `dilution`.)

`time` (Numeric) Incubation time (in hours) - from the start of incubation to when the sample measurements were taken. (Defaults to `NULL`.) (Note: Single entry only, use only if all samples were taken after the same amount of incubation time.)

`time.col` (Character) Column name containing `time` information. (Defaults to "Time".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `time`.)

`istd.col` (Character) Column name of `data.in` containing the MS peak area for the internal standard. (Defaults to "ISTD.Area".)

`istd.name` (Character) The identity of the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds use the same internal standard.)

`istd.name.col` (Character) Column name containing `istd.name` information. (Defaults to "ISTD.Name".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `istd.name`.)

`istd.conc` (Numeric) The concentration for the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds have the same internal standard concentration.)

istd.conc.col (Character) Column name containing istd.conc information. (Defaults to "ISTD.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in istd.conc.)

test.nominal.conc (Numeric) The nominal concentration added to the RED assay at time 0. (Defaults to NULL.) (Note: Single entry only, use only if all tested compounds used the same concentration at time 0.)

test.nominal.conc.col (Character) Column name containing test.nominal.conc information. (Defaults to "Test.Target.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in test.nominal.conc.)

plasma.percent (Numeric) The percent of the physiological plasma concentration used in RED assay. (Defaults to NULL.) (Note: Single entry only, use only if all compounds were tested with the same plasma percent.)

plasma.percent.col (Character) Column name containing plasma.percent information. (Defaults to "Plasma.Percent".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in plasma.percent.)

test.conc (Numeric) The standard test chemical concentration for the fup RED assay. (Defaults to NULL.) (Note: Single entry only, use only if the same standard concentration was used for all tested compounds.)

test.conc.col (Character) Column name containing test.conc information. (Defaults to "Test.Compound.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in test.conc.)

area.col (Character) Column name of data.in containing the target analyte (that is, the test compound) MS peak area. (Defaults to "Area".)

biological.replicates (Character) Replicates with the same analyte. Typically, this uses numbers or letters to index. (Defaults to NULL.) (Note: Single entry only, use only if none of the test compounds have replicates.)

biological.replicates.col (Character) Column name of data.in containing the number or the indices of replicates with the same analyte. (Defaults to "Biological.Replicates".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in biological.replicates.)

technical.replicates (Character) Repeated measurements from one sample. Typically, this uses numbers or letters to index. (Defaults to NULL.) (Note: Single entry only, use only if none of the test compounds have replicates.)

technical.replicates.col (Character) Column name of data.in containing the number or the indices of replicates taken from the one sample. (Defaults to "Technical.Replicates".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in technical.replicates.)

`analysis.method`
: (Character) The analytical chemistry analysis method, typically "LCMS" or "GCMS", liquid chromatography or gas chromatography–mass spectrometry, respectively. (Defaults to `NULL`.) (Note: Single entry only, use only if the same method was used for all tested compounds.)

`analysis.method.col`
: (Character) Column name containing `analysis.method` information. (Defaults to "Analysis.Method".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.method`.)

`analysis.instrument`
: (Character) The instrument used for chemical analysis, for example "Waters ACQUITY I-Class UHPLC - Xevo TQ-S uTQMS". (Defaults to `NULL`.) (Note: Single entry only, use only if the same instrument was used for all tested compounds.)

`analysis.instrument.col`
: (Character) Column name containing `analysis.instrument` information. (Defaults to "Analysis.Instrument".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.instrument`.)

`analysis.parameters`
: (Character) The parameters used to identify the compound on the chemical analysis instrument. (Defaults to `NULL`.) (Note: Single entry only, use only if the same parameters were used for all tested compounds.)

`analysis.parameters.col`
: (Character) Column name containing `analysis.parameters` information. (Defaults to "Analysis.Parameters".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `analysis.parameters`.)

`note.col`
: (Character) Column name of `data.in` containing additional notes on test compounds. (Defaults to "Note".)

`level0.file`
: (Character) The level-0 file from which the `data.in` were obtained. (Defaults to `NULL`.) (Note: Single entry only, use only if all rows in `data.in` were obtained from the same level-0 file.)

`level0.file.col`
: (Character) Column name containing `level0.file` information. (Defaults to "Level0.File".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.file`.)

`level0.sheet`
: (Character) The specific sheet name of level-0 file from which the `data.in` is obtained from, if the level-0 file is an Excel workbook. (Defaults to `NULL`.) (Note: Single entry only, use only if all rows in `data.in` were obtained from the same sheet in the same level-0 file.)

`level0.sheet.col`
: (Character) Column name containing `level0.sheet` information. (Defaults to "Level0.Sheet".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.sheet`.)

output.res	(Logical) When set to TRUE, the result table (level-1) will be exported to the user's per-session temporary directory or OUTPUT.DIR (if specified) as a .tsv file. (Defaults to FALSE.)
save.bad.types	(Logical) When set to TRUE, export data removed due to inappropriate sample types. See the Detail section for the required sample types. (Defaults to FALSE.)
sig.figs	(Numeric) The number of significant figures to round the exported result table (level-1). (Defaults to 5.)
INPUT.DIR	(Character) Path to the directory where the input level-0 file exists. If NULL, looking for the input level-0 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The data frame of observations should be annotated according to these types:

No Plasma Blank (no chemical, no plasma)	NoPlasma.Blank
Plasma Blank (no chemical, just plasma)	Plasma.Blank
Plasma well concentration	Plasma
Phosphate-buffered well concentration	PBS
Time zero plasma concentration	T0
Plasma stability sample	Stability
Acceptor compartment of the equilibrium evaluation	EC_acceptor
Donor compartment of the equilibrium evaluation (chemical spiked side)	EC_donor
Calibration Curve	CC

Chemical concentration is calculated qualitatively as a response and returned as a column in the output data frame:

Response <- AREA / ISTD.AREA * ISTD.CONC

If the output level-1 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, INPUT.DIR and/or OUTPUT.DIR should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-1 data frame with a standardized format containing a standardized set of columns and column names with plasma protein binding (PPB) data from an rapid equilibrium dialysis (RED) assay.

Author(s)

John Wambaugh

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Examples

```
## Load the example level-0 data
level0 <- invitroTKstats::fup_red_L0

## Run it through level-1 processing function
## This example shows the use of the data.in argument which allows users to pass
## in a data frame from the R session.
## If the input level-0 data exists in an external file such as a .tsv file,
## users may import it using FILENAME and INPUT.DIR to specify the file name
## and its directory path, respectively.
level1 <- format_fup_red(data.in = level0,
                         sample.col ="Sample",
                         date.col="Date",
                         compound.col="Compound",
                         lab.compound.col="Lab.Compound.ID",
                         type.col="Sample.Type",
                         dilution.col="Dilution.Factor",
                         technical.replicates.col ="Replicate",
                         biological.replicates = 1,
                         cal=1,
                         area.col = "Peak.Area",
                         istd.conc = 10/1000,
                         istd.col= "ISTD.Peak.Area",
                         test.conc.col = "Compound.Conc",
                         test.nominal.conc = 10,
                         plasma.percent = 100,
                         time.col = "Time",
                         analysis.method = "LCMS",
                    analysis.instrument = "Waters ACQUITY I-Class UHPLC - Xevo TQ-S uTQMS",
                         analysis.parameters = "RT",
                         note.col=NULL,
                         output.res = FALSE
                         )
```

`format_fup_uc` *Creates a Standardized Data Frame with Ultracentrifugation (UC) Plasma Protein Binding (PPB) Data (Level-1)*

Description

This function formats data describing mass spectrometry (MS) peak areas from samples collected as part of *in vitro* measurements of chemical fraction unbound in plasma using ultracentrifugation (Redgrave et al. 1975). The input data frame is organized into a standard set of columns and written to a tab-separated text file.

Usage

```
format_fup_uc(
  FILENAME = "MYDATA",
  data.in,
  sample.col = "Lab.Sample.Name",
  lab.compound.col = "Lab.Compound.Name",
  dtxsid.col = "DTXSID",
  date = NULL,
  date.col = "Date",
  compound.col = "Compound.Name",
  area.col = "Area",
  type.col = "Sample.Type",
  test.conc = NULL,
  test.conc.col = "Test.Compound.Conc",
  cal = NULL,
  cal.col = "Cal",
  dilution = NULL,
  dilution.col = "Dilution.Factor",
  istd.col = "ISTD.Area",
  istd.name = NULL,
  istd.name.col = "ISTD.Name",
  istd.conc = NULL,
  istd.conc.col = "ISTD.Conc",
  test.nominal.conc = NULL,
  test.nominal.conc.col = "Test.Target.Conc",
  biological.replicates = NULL,
  biological.replicates.col = "Biological.Replicates",
  technical.replicates = NULL,
  technical.replicates.col = "Technical.Replicates",
  analysis.method = NULL,
  analysis.method.col = "Analysis.Method",
  analysis.instrument = NULL,
  analysis.instrument.col = "Analysis.Instrument",
  analysis.parameters = NULL,
  analysis.parameters.col = "Analysis.Parameters",
  note.col = "Note",
  level0.file = NULL,
  level0.file.col = "Level0.File",
  level0.sheet = NULL,
  level0.sheet.col = "Level0.Sheet",
  output.res = FALSE,
```

```
  save.bad.types = FALSE,
  sig.figs = 5,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the output level-1 file, "<FILENAME>-fup-UC-Level1.tsv", and/or used to identify the input level-0 file, "<FILENAME>-fup-UC-Level0.tsv" if importing from a .tsv file. (Defaults to "MYDATA".)
data.in	(Data Frame) A level-0 data frame containing mass-spectrometry peak areas, indication of chemical identity, and measurement type. The data frame should contain columns with names specified by the following arguments:
sample.col	(Character) Column name from data.in containing the unique mass spectrometry (MS) sample name used by the laboratory. (Defaults to "Lab.Sample.Name".)
lab.compound.col	(Character) Column name from data.in containing the test compound name used by the laboratory. (Defaults to "Lab.Compound.Name".)
dtxsid.col	(Character) Column name from data.in containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard). (Defaults to "DTXSID".)
date	(Character) The laboratory measurement date, format "MMDDYY" where "MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year. (Defaults to NULL.) (Note: Single entry only, use only if all data were collected on the same date.)
date.col	(Character) Column name containing date information. (Defaults to "Date".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in date.)
compound.col	(Character) Column name from data.in containing the test compound. (Defaults to "Compound.Name".)
area.col	(Character) Column name from data.in containing the target analyte (that is, the test compound) MS peak area. (Defaults to "Area".)
type.col	(Character) Column name from data.in containing the sample type (see table under Details). (Defaults to "Sample.Type".)
test.conc	(Numeric) The standard test chemical concentration for the fup UC assay. (Defaults to NULL.) (Note: Single entry only, use only if the same standard concentration was used for all tested compounds.)
test.conc.col	(Character) Column name containing test.conc information. (Defaults to Test.Compound.Conc".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in test.conc.)
cal	(Character) MS calibration the samples were based on. Typically, this uses indices or dates to represent if the analyses were done on different machines on the same day or on different days with the same MS analyzer. (Defaults to NULL.) (Note: Single entry only, use only if all data were collected based on the same calibration.)

`cal.col` (Character) Column name containing `cal` information. (Defaults to "Cal".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `cal`.)

`dilution` (Numeric) Number of times the sample was diluted before MS analysis. (Defaults to `NULL`.) (Note: Single entry only, use only if all samples underwent the same number of dilutions.)

`dilution.col` (Character) Column name containing `dilution` information. (Defaults to "Dilution.Factor".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `dilution`.)

`istd.col` (Character) Column name of `data.in` containing the MS peak area for the internal standard. (Defaults to "ISTD.Area".)

`istd.name` (Character) The identity of the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds use the same internal standard.)

`istd.name.col` (Character) Column name containing `istd.name` information. (Defaults to "ISTD.Name".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `istd.name`.)

`istd.conc` (Numeric) The concentration for the internal standard. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds have the same internal standard concentration.)

`istd.conc.col` (Character) Column name containing `istd.conc` information. (Defaults to "ISTD.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `istd.conc`.)

`test.nominal.conc`
(Numeric) The nominal concentration added to the UC assay at time 0. (Defaults to `NULL`.) (Note: Single entry only, use only if all tested compounds used the same concentration at time 0.)

`test.nominal.conc.col`
(Character) Column name containing `test.nominal.conc` information. (Defaults to "Test.Target.Conc".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `test.nominal.conc`.)

`biological.replicates`
(Character) Replicates with the same analyte. Typically, this uses numbers or letters to index. (Defaults to `NULL`.) (Note: Single entry only, use only if none of the test compounds have replicates.)

`biological.replicates.col`
(Character) Column name of `data.in` containing the number or the indices of replicates with the same analyte. (Defaults to "Biological.Replicates".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `biological.replicates`.)

`technical.replicates`
(Character) Repeated measurements from one sample. Typically, this uses numbers or letters to index. (Defaults to `NULL`.) (Note: Single entry only, use only if none of the test compounds have replicates.)

technical.replicates.col
: (Character) Column name of data.in containing the number or the indices of replicates taken from the one sample. (Defaults to "Technical.Replicates".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in technical.replicates.)

analysis.method
: (Character) The analytical chemistry analysis method, typically "LCMS" or "GCMS", liquid chromatography or gas chromatography–mass spectrometry, respectively. (Defaults to NULL.) (Note: Single entry only, use only if the same method was used for all tested compounds.)

analysis.method.col
: (Character) Column name containing analysis.method information. (Defaults to "Analysis.Method".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.method.)

analysis.instrument
: (Character) The instrument used for chemical analysis, for example "Waters Xevo TQ-S micro (QEB0036)". (Defaults to NULL.) (Note: Single entry only, use only if the same instrument was used for all tested compounds.)

analysis.instrument.col
: (Character) Column name containing analysis.instrument information. (Defaults to "Analysis.Instrument".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.instrument.)

analysis.parameters
: (Character) The parameters used to identify the compound on the chemical analysis instrument. (Defaults to NULL.) (Note: Single entry only, use only if the same parameters were used for all tested compounds.)

analysis.parameters.col
: (Character) Column name containing analysis.parameters information. (Defaults to "Analysis.Parameters".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in analysis.parameters.)

note.col
: (Character) Column name of data.in containing additional notes on the test compounds. (Defaults to "Note").

level0.file
: (Character) The level-0 file from which the data.in were obtained. (Defaults to NULL.) (Note: Single entry only, use only if all rows in data.in were obtained from the same level-0 file.)

level0.file.col
: (Character) Column name containing level0.file information. (Defaults to "Level0.File".) (Note: data.in does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in level0.file.)

level0.sheet
: (Character) The specific sheet name of the level-0 file where data.in is obtained from, if the level-0 file is an Excel workbook. (Defaults to NULL.) (Note: Single entry only, use only if all rows in data.in were obtained from the same sheet in the same level-0 file.)

`level0.sheet.col`
: (Character) Column name containing `level0.sheet` information. (Defaults to "Level0.Sheet".) (Note: `data.in` does not necessarily have this field. If this field is missing, it can be auto-filled with the value specified in `level0.sheet`.)

`output.res`
: (Logical) When set to `TRUE`, the result table (level-1) will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)

`save.bad.types`
: (Logical) When set to `TRUE`, export data removed due to inappropriate sample types. See the Detail section for the required sample types. (Defaults to `FALSE`.)

`sig.figs`
: (Numeric) The number of significant figures to round the exported result table (level-1). (Defaults to 5.)

`INPUT.DIR`
: (Character) Path to the directory where the input level-0 file exists. If `NULL`, looking for the input level-0 file in the current working directory. (Defaults to `NULL`.)

`OUTPUT.DIR`
: (Character) Path to the directory to save the output file. If `NULL`, the output file will be saved to the user's per-session temporary directory or `INPUT.DIR` if specified. (Defaults to `NULL`.)

`verbose`
: (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The data frame of observations should be annotated according to these types:

Calibration Curve	CC
Ultracentrifugation Aqueous Fraction	AF
Whole Plasma T1h Sample	T1
Whole Plasma T5h Sample	T5

Chemical concentration is calculated qualitatively as a response and returned as a column in the output data frame:

Response <- AREA / ISTD.AREA * ISTD.CONC

If the output level-1 result table is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-1 data frame with a standardized format containing a standardized set of columns and column names with plasma protein binding (PPB) data from an ultracentrifugation (UC) assay.

Author(s)

John Wambaugh

References

Redgrave TG, Roberts DCK, West CE (1975). "Separation of plasma lipoproteins by density-gradient ultracentrifugation." *Analytical Biochemistry*, **65**(1–2), 42–49.

Examples

```
## Load the example level-0 data
level0 <- invitroTKstats::fup_uc_L0

## Run it through level-1 processing function
## This example shows the use of data.in argument which allows users to pass
## in a data frame from the R session.
## If the input level-0 data exists in an external file such as a .tsv file,
## users may import it using INPUT.DIR to specify the path and FILENAME
## to specify the file name. See documentation for details.
level1 <- format_fup_uc(data.in = level0,
                        sample.col="Sample",
                        compound.col="Compound",
                        test.conc.col ="Compound.Conc",
                        lab.compound.col="Lab.Compound.ID",
                        type.col="Sample.Type",
                        istd.col="ISTD.Peak.Area",
                        cal.col = "Date",
                        area.col = "Peak.Area",
                        istd.conc = 1,
                        note.col = NULL,
                        test.nominal.conc = 10,
                        analysis.method = "UPLC-MS/MS",
                        analysis.instrument = "Waters Xevo TQ-S micro (QEB0036)",
                        analysis.parameters.col = "Analysis.Params",
                        technical.replicates.col = "Replicate",
                        biological.replicates = 1,
                        output.res = FALSE
                        )
```

---

fup_red_cheminfo *Fup RED Chemical Information Example Data set*

Description

The chemical ID mapping information from mass spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set contains 26 unique compounds/chemicals.

Usage

```
fup_red_cheminfo
```

Format

A chemical info data.frame with 26 rows and 4 variables:

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`NAME (Abbreviation)` Name of the test analyte/compound and abbreviation used by the lab as the compound ID

`Compound` Name of the test analyte/compound

`Chem.Lab.ID` Abbreviation of the test analyte/compound as described in the laboratory

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

---

fup_red_L0 *Fup RED Level-0 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_red_L0
```

Format

A level-0 data.frame with 660 rows and 18 variables:

`Compound` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.ID` Compound as described in the laboratory

`Date` Date the sample was added to the MS analyzer

`Sample` Sample description used in the laboratory

`Type` Type of RED sample, annotated by the laboratory

`Compound.Conc` Expected (or nominal) concentration of analyte (for calibration curve)

`Peak.Area` Peak area of analyte (target compound)

`ISTD.Peak.Area` Peak area of internal standard (ISTD) compound (pixels)

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`Analysis.Params` Column contains the retention time

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Sample Text` Additional notes on the sample

`Sample.Type` Type of RED sample in `invitroTKstats` package annotations

`Replicate` Identifier for repeated measurements of one sample of a compound

`Time` Time when the sample was measured - in hours (h)

`Dilution.Factor` Number of times the sample was diluted

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_red_L1` *Fup RED Level-1 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_red_L1
```

Format

A level-1 data.frame with 636 rows and 25 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of RED sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to RED plate (uM)

`Percent.Physiologic.Plasma` Percent of physiological plasma concentration in RED plate (in percent)

`Technical.Replicates` Identifier for repeated measurements of a sample of a compound

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_red_L2` *Fup RED Level-2 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_red_L2
```

Format

A level-2 data.frame with 636 rows and 26 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of RED sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to RED plate (uM)

`Percent.Physiologic.Plasma` Percent of physiological plasma concentration in RED plate (in percent)

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If, "Y" then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_red_L2_heldout` *Fup RED Level-2 Heldout Example Data set*

Description

The unverified level-2 samples from mass spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 0 test analytes/compounds. No data samples are unverified.

Usage

```
fup_red_L2_heldout
```

Format

A level-2 data.frame with 0 rows and 26 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of RED sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Time` Time when the sample was measured - in hours (h)

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to RED plate (uM)

`Percent.Physiologic.Plasma` Percent of physiological plasma concentration in RED plate (in percent)

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_red_L3` *Fup RED Level-3 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_red_L3
```

Format

A level-3 data.frame with 3 rows and 4 variables:

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Calibration` Identifier for mass spectrometry calibration – usually the date

`Fup` Fraction unbound in plasma

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

---

fup_red_L4 *Fup RED Level-4 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_red_L4
```

Format

A level-4 data.frame with 3 rows and 7 variables:

`Compound.Name` Name of the test analyte/compound

`Lab.Compound.Name` Compound as described in the laboratory

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Fup.point` Point estimate of fraction unbound in plasma

`Fup.Med` Median fraction unbound in plasma

`Fup.Low` 2.5th quantile of fraction unbound in plasma

`Fup.High` 97.5th quantile of fraction unbound in plasma

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_red_PREJAGS` *Fup RED Level-4 PREJAGS arguments*

Description

The arguments given to JAGS for the tested compound during level-4 processing of mass spectrometry measurements of plasma protein binding (PPB) via rapid equilibrium dialysis (RED) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This list is overwritten for each tested compound. Therefore, only contains arguments given to JAGS for the last tested compound.

Usage

```
fup_red_PREJAGS
```

Format

A named list with 33 elements:

`Test.Nominal.Conc` Unique `Test.Nominal.Conc` values (expected initial concentration) for the tested compound

`Num.cal` Unique number of `Calibration` values for the tested compound

`Physiological.Protein.Conc` The assumed physiological protein concentration for plasma protein binding calculations. (Defaults to 70/(66.5*1000)*1000000. According to Berg and Lane (2011): 60-80 mg/mL, albumin is 66.5 kDa, assume all protein is albumin to estimate default in uM.)

`Assay.Protein.Perecent` `Percent.Physiologic.Plasma` values for each "Plasma" sample type replicate group

`Num.Plasma.Blank.obs` Number of "Plasma.Blank" sample types for the tested compound

`Plasma.Blank.obs` `Response` of the "Plasma.Blank" sample types for the tested compound

`Plasma.Blank.cal` Indices of the unique `Calibration` values that corresponds to the "Plasma.Blank" sample types' `Calibration` for the tested compound

`Plasma.Blank.df` Unique `Dilution` `Factor` of the "Plasma.Blank" sample types for the tested compound

`Plasma.Blank.rep` Integer representing "Plasma.Blank" replicate group for the tested compound

`Num.NoPlasma.Blank.obs` Number of "NoPlasma.Blank" sample types for the tested compound

NoPlasma.Blank.obs Response of the "NoPlasma.Blank" sample types for the tested compound

NoPlasma.Blank.cal Indices of the unique Calibration values that corresponds to the "NoPlasma.Blank" sample types' Calibration for the tested compound

NoPlasma.Blank.df Unique Dilution Factor of the "NoPlasma.Blank" sample types for the tested compound

Num.CC.obs Number of "CC" sample types with non-NA Test.Compound.Conc values for the tested compound

CC.conc Test.Compound.Conc (non-NA) of the "CC" sample types for the tested compound

CC.obs Response of the "CC" sample types with non-NA Test.Compound.Conc for the tested compound

CC.cal Indices of the unique Calibration values that corresponds to the "CC" sample types' Calibration for the tested compound

CC.df Unique Dilution Factor of the "NoPlasma.Blank" sample types for the tested compound

Num.T0.obs Number of "T0" sample types for the tested compound

T0.obs Response of the "T0" sample types for the tested compound

T0.cal Indices of the unique Calibration values that corresponds to the "T0" sample types' Calibration for the tested compound

T0.df Unique Dilution Factor of the "T0" sample types for the tested compound

Num.rep Unique number of (Calibration + Technical.Replicates) combinations for "PBS" and "Plasma" sample types for the tested compound

Num.PBS.obs Number of "PBS" sample types for the tested compound

PBS.obs Response of the "PBS" sample types for the tested compound

PBS.cal Indices of the unique Calibration values that corresponds to the "PBS" sample types' Calibration for the tested compound

PBS.df Unique Dilution Factor of the "PBS" sample types for the tested compound

PBS.rep Integer representing "PBS" replicate group for the tested compound

Num.Plasma.obs Number of "Plasma" sample types for the tested compound

Plasma.obs Response of the "Plasma" sample types for the tested compound

Plasma.cal Indices of the unique Calibration values that corresponds to the "Plasma" sample types' Calibration for the tested compound

Plasma.df Unique Dilution Factor of the "Plasma" sample types for the tested compound

Plasma.rep Integer representing "Plasma" replicate group for the tested compound

References

Waters NJ, Jones R, Williams G, Sohal B (2008). "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences*, **97**(10), 4586–4595.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

fup_uc_cheminfo *Fup UC Chemical Information Example Data set*

Description

The chemical ID mapping information from mass spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set contains 75 unique compounds/chemicals.

Usage

```
fup_uc_cheminfo
```

Format

A chemical info data.frame with 75 rows and 4 variables:

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Chemical Name (Common Abbreviation)` Name of the test analyte/compound and abbreviation used by the lab as the compound ID

`Compound` Name of the test analyte/compound

`Chem.Lab.ID` Common abbreviation of the test analyte/compound as described in the laboratory

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

fup_uc_L0 *Fup UC Level-0 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_uc_L0
```

Format

A level-0 data.frame with 240 rows and 17 variables:

`Compound` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.ID` Compound as described in the laboratory

`Date` Date the sample was added to the MS analyzer

`Sample` Sample description used in the laboratory

`Type` Type of UC sample, annotated by the laboratory

`Compound.Conc` Expected (or nominal) concentration of analyte (for calibration curve)

`Peak.Area` Peak area of analyte (target compound)

`ISTD.Peak.Area` Peak area of internal standard (ISTD) compound (pixels)

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`Analysis.Params` Column contains the retention time

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Sample.Text` Additional notes on the sample

`Sample.Type` Type of UC sample in `invitroTKstats` package annotations

`Dilution.Factor` Number of times the sample was diluted

`Replicate` Identifier for repeated measurements of one sample of a compound

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

---

fup_uc_L1 *Fup UC Level-1 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_uc_L1
```

Format

A level-1 data.frame with 240 rows and 23 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of UC sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to T1 sample (uM)

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Response` Response factor (calculated from analyte and ISTD peaks)

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

fup_uc_L2 *Fup UC Level-2 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_uc_L2
```

Format

A level-2 data.frame with 240 rows and 24 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of UC sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to T1 sample (uM)

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_uc_L2_heldout` *Fup UC Level-2 Heldout Example Data set*

Description

The unverified level-2 samples from mass spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 0 test analytes/compounds. No data samples are unverified.

Usage

```
fup_uc_L2_heldout
```

Format

A level-2 data.frame with 0 rows and 24 variables:

`Lab.Sample.Name` Sample description used in the laboratory

`Date` Date the sample was added to the MS analyzer

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Sample.Type` Type of UC sample

`Dilution.Factor` Number of times the sample was diluted

`Calibration` Identifier for mass spectrometry calibration – usually the date

`ISTD.Name` Name of the internal standard (ISTD) analyte/compound

`ISTD.Conc` Concentration of ISTD (uM)

`ISTD.Area` Peak area of internal standard (ISTD) compound (pixels)

`Area` Peak area of analyte (target compound)

`Analysis.Method` General description of chemical analysis method

`Analysis.Instrument` Instrument(s) used for chemical analysis

`Analysis.Parameters` Parameters for identifing analyte peak (for example, retention time)

`Note` Any laboratory notes about sample

`Level0.File` Name of the laboratory data file from which the level-0 sample data was extracted

`Level0.Sheet` Name of the Excel workbook 'sheet' from which the level-0 sample data was extracted

`Test.Compound.Conc` Measured concentration of analytic standard (for calibration curve) (uM)

`Test.Nominal.Conc` Expected initial concentration of chemical added to T1 sample (uM)

`Biological.Replicates` Identifier for measurements of multiple samples with the same analyte

`Technical.Replicates` Identifier for repeated measurements of one sample of a compound

`Response` Response factor (calculated from analyte and ISTD peaks)

`Verified` If "Y", then sample is included in the analysis. (Any other value causes the data to be ignored.)

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_uc_L3` *Fup UC Level-3 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_uc_L3
```

Format

A level-3 data.frame with 3 rows and 5 variables:

`Compound.Name` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Calibration` Identifier for mass spectrometry calibration – usually the date

`Fup` Fraction unbound in plasma

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

`fup_uc_L4` *Fup UC Level-4 Example Data set*

Description

Mass Spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This data set is a subset of experimental data containing samples for 3 test analytes/compounds.

Usage

```
fup_uc_L4
```

Format

A level-4 data.frame with 3 rows and 10 variables:

`Compound` Name of the test analyte/compound

`DTXSID` DSSTox Substance Identifier (CompTox Chemicals Dashboard - CCD)

`Lab.Compound.Name` Compound as described in the laboratory

`Fstable.Med` Median stability fraction

`Fstable.Low` 2.5th quantile of stability fraction

`Fstable.High` 97.5th quantile of stability fraction

`Fup.Med` Median fraction unbound in plasma

`Fup.Low` 2.5th quantile of fraction unbound in plasma

`Fup.High` 97.5th quantile of fraction unbound in plasma

`Fup.point` Point estimate of fraction unbound in plasma

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

fup_uc_PREJAGS *Fup UC Level-4 PREJAGS arguments*

Description

The arguments given to JAGS for the tested compound during level-4 processing of mass spectrometry measurements of plasma protein binding (PPB) via ultracentrifugation (UC) for per- and poly-fluorinated alkyl substance (PFAS) samples. Experiments were led by Dr.s Marci Smeltz and Barbara Wetmore (Smeltz et al. 2023). This list is overwritten for each tested compound. Therefore, only contains arguments given to JAGS for the last tested compound.

Usage

```
fup_uc_PREJAGS
```

Format

A named list with 10 elements:

`Num.cal` Unique number of `Calibration` values for the tested compound

`Num.obs` Total number of observations for the tested compound

`Response.obs` `Response` of all samples for the tested compound

`obs.conc` Indices of the `Test.Compound.Conc` values that corresponds to all samples' `Test.Compound.Conc` for the tested compound.

`obs.cal` Indices of the unique `Calibration` values that corresponds to all samples' `Calibration` for the tested compound.

`Conc` `Test.Compound.Conc` of the "CC" sample types + three placeholder concentrations ("T1", "T5", "AF") per `Biological.Replicates` series

`Num.cc.obs` Number of "CC" sample types for the tested compound

`Num.series` Unique number of `Biological.Replicates` series

`Dilution.Factor` `Dilution.Factor` of all samples for the tested compound (number of times the sample was diluted)

`Test.Nominal.Conc` Unique `Test.Nominal.Conc` values (expected initial concentration) of all samples for the tested compound

References

Howard ML, Hill JJ, Galluppi GR, McLean MA (2010). "Plasma protein binding in drug discovery and development." *Combinatorial chemistry & high throughput screening*, **13**(2), 170–187.

Smeltz M, Wambaugh JF, Wetmore BA (2023). "Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment." *Chemical Research in Toxicology*, **36**(6), 870–881.

Heaviside *Heaviside*

Description

Evaluate the Heaviside function with `threshold` indicating the discontinuity. If elements in `x` are greater than or equal to `threshold`, returns 1. Otherwise, returns 0.

Usage

```
Heaviside(x, threshold = 0)
```

Arguments

`x`	(Numeric) A numeric vector.
`threshold`	(Numeric) A threshold value used to compare to elements in `x`. (Defaults to 0.)

Value

A vector of 1 and 0. 1 indicates the element in `x` is larger or equal to the `threshold`.

initfunction_clint *Set Initial Values for Intrinsic Hepatic Clearance (Clint) Bayesian Model*

Description

Sets the initial values of arguments required for JAGS such as assumed initial probability distributions. The list is used as an argument to JAGS during level-4 processing.

Usage

```
initfunction_clint(mydata, seed)
```

Arguments

`mydata`	(List) Output of `build_mydata_clint`.
`seed`	(Numeric) Random Number Generator (RNG) seed to use for reproducibility.

Value

A list of initial values.

initfunction_fup_red *Set Initial Values for Fup RED Bayesian Model*

Description

Sets the initial values of arguments required for JAGS such as assumed initial probability distributions. The list is used as an argument to JAGS during level-4 processing.

Usage

```
initfunction_fup_red(mydata, seed)
```

Arguments

mydata	(List) Output of build_mydata_fup_red.
seed	(Numeric) Random Number Generator (RNG) seed to use for reproducibility.

Value

A list of initial values.

initfunction_fup_uc *Set Initial Values for Fup UC Bayesian Model*

Description

Sets the initial values of arguments required for JAGS such as assumed initial probability distributions. The list is used as an argument to JAGS during level-4 processing.

Usage

```
initfunction_fup_uc(mydata, seed)
```

Arguments

mydata	(List) Output of build_mydata_fup_uc.
seed	(Numeric) Random Number Generator (RNG) seed to use for reproducibility.

Value

A list of initial values.

L1.common.col *Common Columns in Level-1*

Description

Common column names across the various *in vitro* assays used for collecting *in vitro* toxicokinetic parameters.

Usage

```
L1.common.cols
```

Format

A named character vector containing the default/standard column names across HTTK assays, where the element names are the corresponding L1 arguments.

merge_level0 *Merge Multiple Level-0 files into a Single Table for Processing*

Description

This function reads multiple Excel files containing mass-spectrometry (MS) data and extracts the chemical sample data from the specified sheets. The argument 'level0.catalog' is a table that provides the necessary information to find the data for each chemical. The primary data of interest are the analyte peak area, the internal standard peak area, and the target concentration for calibration curve (CC) samples. The argument 'data.label' is used to annotate this particular mapping of level-0 files into data ready to be organized into a level-1 file.

Usage

```
merge_level0(
  FILENAME = "MYDATA",
  level0.catalog,
  file.col = "File",
  sheet = NULL,
  sheet.col = "Sheet",
  skip.rows = NULL,
  skip.rows.col = "Skip.Rows",
  num.rows = NULL,
  num.rows.col = NULL,
  date = NULL,
  date.col = "Date",
  compound.col = "Chemical.ID",
  istd.col = "ISTD",
```

```
  col.names.loc = NULL,
  col.names.loc.col = "Col.Names.Loc",
  sample.colname = NULL,
  sample.colname.col = "Sample.ColName",
  type.colname = NULL,
  type.colname.col = "Type",
  peak.colname = NULL,
  peak.colname.col = "Peak.ColName",
  istd.peak.colname = NULL,
  istd.peak.colname.col = "ISTD.Peak.ColName",
  conc.colname = NULL,
  conc.colname.col = "Conc.ColName",
  analysis.param.colname = NULL,
  analysis.param.colname.col = "AnalysisParam.ColName",
  additional.colnames = NULL,
  additional.colname.cols = NULL,
  chem.ids,
  chem.lab.id.col = "Chem.Lab.ID",
  chem.name.col = "Compound",
  chem.dtxsid.col = "DTXSID",
  catalog.out = FALSE,
  output.res = FALSE,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify outputs of the function call. (Default to "MYDATA")
level0.catalog	A data frame describing which columns of which sheets in which Excel files contain MS data for analysis. See details for full explanation.
file.col	(Character) Column name containing level-0 file names to pull data from.
sheet	(Character) Excel file sheet name/identifier containing level-0 where data is to be pulled from. (Defaults to 'NULL'.) (Note: Single entry only, use only if all files have the same sheet identifier for level-0 data.)
sheet.col	(Character) Catalog column name containing 'sheet' information. (Default to "Sheet")
skip.rows	(Numeric) Number of rows to skip when extracting level-0 data from the specified Excel file(s). (Defaults to 'NULL'.) (Note: Single entry only, use only if all files need to skip the same number of rows for extracting level-0 data.)
skip.rows.col	(Character) Catalog column name containing 'skip.rows' information. (Default to "Skip.Rows")
num.rows	(Numeric) Number of rows to pull when extracting level-0 data from the specified Excel file(s). (Defaults to 'NULL'.) (Note: Single entry only, use only if all files need to pull the same number of rows for extracting level-0 data.)

`num.rows.col` (Character) Catalog column name containing ‘num.rows‘ information. (Default to ‘NULL‘)

`date` (Character) Date of laboratory measurements. Typical format "MMDDYY" ("MM" = 2 digit month, "DD" = 2 digit day, and "YY" = 2 digit year). (Defaults to ‘NULL‘.) (Note: Single entry only, use only if all files have the same laboratory measurement date.)

`date.col` (Character) Catalog column name containing ‘date‘ information. (Defaults to "Date")

`compound.col` (Character) Catalog column name containing ‘compound‘ information. (Defaults to "Chemical.ID")

`istd.col` (Character) Catalog column name containing ‘istd‘ information, or the MS peak area for the internal standard. (Defaults to "ISTD")

`col.names.loc` (Numeric) Row location of data column names. (Defaults to ’NULL’.) (Note: Single entry only, use only if all files have column names in the same row location, typically the first row.)

`col.names.loc.col` (Character) Catalog column name containing ‘col.names.loc‘ information. (Defaults to "Col.Names.Loc")

`sample.colname` (Character) Column name of level-0 data containing sample information. (Defaults to ‘NULL‘.) (Note: Single entry only, use only if all files use the same column name for sample names when extracting level-0 data.)

`sample.colname.col` (Character) Catalog column name containing ‘sample.colname‘ information. (Defaults to "Sample.ColName")

`type.colname` (Character) Column name of the level-0 data containing the type of sample. (Defaults to ‘NULL‘.) (Note: Single entry only, use only if all files use the same column name for sample type information when extracting level-0 data.)

`type.colname.col` (Character) Catalog column name containing ‘type.colname‘ information. (Defaults to "Type".)

`peak.colname` (Character) Column name of the level-0 data containing the analyte Mass Spectrometry peak area. (Defaults to ‘NULL‘.) (Note: Single entry only, use only if all files use the same column name for analyte peak area information when extracting level-0 data.)

`peak.colname.col` (Character) Catalog column name containing ‘peak.colname‘ information. (Defaults to "Peak.ColName")

`istd.peak.colname` (Character) Column name of the level-0 data containing the internal standard Mass Spectrometry peak area. (Note: Single entry only, use only if all files use the same column name for internal standard MS peak area information when extracting level-0 data.)

`istd.peak.colname.col` (Character) Catalog column name containing ‘istd.peak.colname‘ information. (Defaults to "ISTD.Peak.ColName")

`conc.colname` (Character) Column name of the level-0 data containing intended concentrations for calibration curves. (Defaults to 'NULL'.) (Note: Single entry only, use only if all files use the same column name for intended concentration information when extracting level-0 data.)

`conc.colname.col` (Character) Catalog column name containing 'conc.colname' information. (Defaults to "Conc.ColName")

`analysis.param.colname` (Character) Column name of the level-0 data containing Mass Spectrometry instrument parameters for the analyte. (Defaults to 'NULL'.) (Note: Single entry only, use only if all files use the same column name for analysis parameter information when extracting level-0 data.)

`analysis.param.colname.col` (Character) Catalog column name containing 'analysis.param.colname' information. (Defaults to "AnalysisParam.ColName")

`additional.colnames` Additional columns from the level-0 data files to pull information from when extracting level-0 data and include in the compiled level-0 returned from 'merge_level0'. (Defaults to 'NULL'.)

`additional.colname.cols` Catalog column name(s) containing 'additional.colnames' information, (Defaults to 'NULL'.)

`chem.ids` (Data frame) A data frame containing basic chemical identification information for tested chemicals.

`chem.lab.id.col` (Character) Column in 'chem.ids' containing the compound/chemical identifier used by the laboratory in level-0 measured data. (Defaults to "Chem.Lab.ID")

`chem.name.col` (Character) 'chem.ids' column name containing the "standard" chemical name to use for annotation of the compiled level-0 returned from 'merge_level0'. (Defaults to "Compound")

`chem.dtxsid.col` (Character) 'chem.ids' column name containing EPA's DSSTox Structure ID (http://comptox.epa.gov/dashboard) (Defaults to "DTXSID")

`catalog.out` (Logical) When set to `TRUE`, the data frame specified in `level0.catalog` will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)

`output.res` (Logical) When set to `TRUE`, the result table (level-0) will be exported to the user's per-session temporary directory or `OUTPUT.DIR` (if specified) as a .tsv file. (Defaults to `FALSE`.)

`INPUT.DIR` (Character) Path to the directory where the Excel files with level-0 data exist. If not specified, looking for the files in the current working directory. (Defaults to `NULL`.)

`OUTPUT.DIR` (Character) Path to the directory to save the output file. If `NULL`, the output file will be saved to the user's per-session temporary directory. (Defaults to `NULL`.)

`verbose` (*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

Unless specified to be a single value for all the files, for example sheet="Data", the argument 'level0.catalog' should be a data frame with the following columns:

File	The Excel filename to be loaded
Sheet	The name of the Sheet to examine within in the Excel file
Skip.Rows	How many rows should be skipped on the sheet to get usable column names
Date	The date the measurements were made
Chemical.ID	The laboratory chemical identity
ISTD	The internal standard used
Col.Names.Loc	The row locations of the column names
Sample.ColName	The column name on the sheet that contains sample identity
Type.ColName	The column name on the sheet that contains the type of sample
Peak.ColName	The column name on the sheet that contains the analyte MS peak area
ISTD.Peak.ColName	The column name on the sheet that contains the internal standard MS peak area
Conc.ColName	The column name on the sheet that contains the intended concentration for calibration curves
AnalysisParam.ColName	The column name on the sheet that contains the MS instrument parameters for the analyte

Columns with names ending in ".ColName" indicate the columns to be extracted from the specified Excel file and sheet containing level-0 data.

If the output level-0 file is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` shoud be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

`data.frame` A data.frame in standardized level-0 format

Author(s)

John Wambaugh

Examples

```
# Create level0.catalog data.frame
# Will need to retrieve "Hep_745_949_959_082421_final.xlsx" file from
# inst/extdata/Kreutz-Clint and save it to desired directory.
# Note XLSX file does not need to be saved to current working directory.
catalog <- create_catalog(file = "Hep_745_949_959_082421_final.xlsx",
                          sheet = "Data063021",
                          skip.rows = 44,
                          num.rows = 30,
                          date = "063021",
                          compound = "745",
```

```
                            istd = "MFBET",
                            sample = "Name",
                            type = "Type",
                            peak = "Area...13",
                            istd.peak = "Resp....16",
                            conc = "Final Conc....11",
                            analysis.param = "RT...12",
                            col.names.loc = 2)
# Create chem.ids data.frame
chem.ids <- data.frame("Chem.Lab.ID" = "745",
                       "Compound" = "(Heptafluorobutanoyl)pivaloylmethane",
                       "DTXSID" = "DTXSID3066215")
# Create level0 data.frame
# Will need to replace <PATH TO FILE> with chosen desired directory containing
# XLSX file from above.
level0 <- merge_level0(level0.catalog = catalog,
             INPUT.DIR = system.file("extdata/Kreutz-Clint",package = "invitroTKstats"),
             istd.col = "ISTD.Name",
             type.colname.col = "Type.ColName",
             num.rows.col = "Number.Data.Rows",
             chem.ids = chem.ids,
             catalog.out = FALSE,
             output.res = FALSE) # do not auto-save the file
```

`plot_clint` *Plot Mass Spectrometry Responses from Measurements of Intrinsic Hepatic Clearance*

Description

This function generates a response-versus-time plot of mass spectrometry (MS) responses collected from measurements of intrinsic hepatic clearance for a chemical. Responses from different measurements/calibrations are labeled with different colors, and responses from various sample types are labeled with different shapes.

Usage

```
plot_clint(level2, dtxsid, color.palette = "viridis")
```

Arguments

`level2`	(Data Frame) A data frame containing level-2 data with a measure of chemical clearance over time when incubated with suspended hepatocytes.
`dtxsid`	(Character) EPA's DSSTox Structure ID for the chemical to be plotted.
`color.palette`	(Character) A character string indicating which `viridis` R package color map option to use. (Defaults to "viridis".)

Details

The function requires "level-2" data for plotting. Level-2 data is level-1, data formatted with the format_clint function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for plotting.

Value

ggplot2	A figure of mass spectrometry responses over time for various sample types.

Author(s)

John Wambaugh

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::clint_L2
plot_clint(level2, dtxsid = "DTXSID1021116")
```

---

plot_fup_uc *Plot Mass Spectrometry Responses for Fraction Unbound in Plasma Data from Ultracentrifugation (UC)*

Description

This function generates a scatter plot of mass spectrometry (MS) responses for one chemical collected from measurement of fraction unbound in plasma (Fup) using ultracentrifugation (UC). The scatter plot displays the MS responses (y-axis) by sample types (x-axis). Responses from different measurements/calibrations are labeled with different shapes and colors.

Usage

```
plot_fup_uc(
  level2,
  dtxsid,
  compare = "type",
  good.col = "Verified",
  color.palette = "viridis"
)
```

Arguments

level2	(Data Frame) A data.frame containing level-2 data for fraction unbound in plasma (Fup) measured by ultracentrifugation (UC).
dtxsid	(Character) EPA's DSSTox Structure ID for the chemical to be plotted.

compare	(Character) A string indicating the plot is for comparing the responses across sample types ("type") or across calibrations ("cal"). (Defaults to "type".)
good.col	(Character) Column name containg verification information, data rows valid for plotting are indicated with a "Y". (Defaults to "Verified".)
color.palette	(Character) A character string indicating which viridis R package color map option to use. (Defaults to "viridis".)

Details

This function requires "level-2" data for plotting. Level-2 data is level-1, data formatted with the format_fup_uc function, and curated with a verification column. "Y" in the verification column indicates the data row is valid for plotting.

Value

ggplot2	A figure of mass spectrometry responses for various sample types.

Author(s)

John Wambaugh

Examples

```
## Load example level-2 data
level2 <- invitroTKstats::fup_uc_L2
plot_fup_uc(level2, dtxsid = "DTXSID0059829")
```

---

round_output *Round Numeric Data (Any Level and Assay)*

Description

This function rounds the numeric columns from any level of processing. Numeric columns may include estimates of chemical-specific toxicokinetic (TK) parameters from the relevant *in vitro* assays or numerical data measurements collected from the mass spectrometry experiments.

Usage

```
round_output(
  FULL_FILENAME = NULL,
  data.in,
  FILENAME = "MYDATA",
  assay = NULL,
  level = NULL,
  exclusion.cols = NULL,
  sig.figs = 3,
```

```
  output.res = FALSE,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FULL_FILENAME	(Character) A string used to identify the full filename of input .tsv or .RData file (i.e. "MYDATA-Clint-Level4.tsv" or "MYDATA-Clint-Level4Analysis-2025-04-23.RData"). The string is also used to name the exported data file (if chosen to be exported). (Note: FULL_FILENAME not required if data.in is provided.) (Defaults to NULL.)
data.in	(Data Frame) Any level data frame generated from invitroTKstats package. (Note: data.in not required if FULL_FILENAME is provided.)
FILENAME	(Character) A string used to name the start of the exported date file. Only required if input data is a data.frame and output file is being exported. (Defaults to "MYDATA".)
assay	(Character) A string used to name the assay used to generate the input data. The string is appended to the name of the exported data file. Only required if input data is a data.frame and output file is being exported. Must be one of the following assays: "Clint", "Caco-2", "fup-RED", or "fup-UC". (Defaults to NULL.)
level	(Character) A string used the name the level of the input data. The string is appended to the name of the exported data file. Only required if input data is a data.frame and output file is being exported. Must be one of the following levels: "0", "1", "2", "3", "4". (Defaults to NULL.)
exclusion.cols	(Character) Vector of column names to exclude from rounding. (Defaults to NULL.)
sig.figs	(Numeric) The number of significant figures to round the desired numeric columns to. (Defaults to 3.)
output.res	(Logical) When set to TRUE, the rounded data file will be exported to the user's per-session temporary directory as a .tsv (if data.in is specified or if FULL_FILENAME is a .tsv) or as an .RData (if FULL_FILENAME is an .RData). (Defaults to FALSE.)
INPUT.DIR	(Character) Path to the directory where the FULL_FILENAME exists. If NULL, looking for the input FULL_FILENAME in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the rounded data file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

For example, for level-3 or level-4 output results, estimates of intrinsic hepatic clearance (Cl~int~) from Hepatocyte Incubation data, fraction unbound in plasma (F~up~) from Rapid Equilibrium

Dialysis (RED) data, fraction unbound in plasma (F~up~) from Ultracentrifugation (UC) data, or apparent membrane permeability from a Caco-2 assay can all be rounded to the desired number of significant figures.

Note: Currently, for level-3 Caco-2 data, the "Frec_A2B.vec" and "Frec_B2A.vec" columns are not rounded. However, these columns can be rounded if the level-3 result table from `calc_caco2_point` is exported and the number of significant figures is specified.

The input to this function can be any level of data (level-0 through level-4) corresponding to any assay (Clint, Caco-2, Fup RED, Fup UC). The desired data object to be rounded can be a data.frame, specified with `data.in`, or a .tsv or .RData, specified with `FULL_FILENAME`.

If the rounded output file is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv or .RData file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A rounded data frame

Author(s)

Lindsay Knupp

Examples

```
## Round Clint-L4 data, exclude p-value columns, and don't export results
level4 <- invitroTKstats::clint_L4
round_output(data.in = level4,
             exclusion.cols = c("Clint.pValue", "Sat.pValue", "degrades.pValue"),
             output.res = FALSE)

## Round Clint-L4 data and export results.
## Note: Will export as a .tsv file.
## Not run:
round_output(data.in = level4, assay = "Clint", level = "4")

## End(Not run)

## Round Clint-L4 .tsv data and export to INPUT.DIR.
## Will need to replace FULL_FILENAME and INPUT.DIR with full filename and location of .tsv.
## Not run:
round_output(FULL_FILENAME = "Example-Clint-Level4.tsv",
             INPUT.DIR = "<FULL_FILENAME FILE LOCATION>")

## End(Not run)

## Round Clint-L4 .RData and export to OUTPUT.DIR
## Will need to replace FULL_FILENAME and INPUT.DIR with full filename and location
```

```
## of .RData. Will also need to replace OUTPUT.DIR with desired location of rounded
## data file.
## Not run:
round_output(FULL_FILENAME = "Example-Clint-Level4Analysis-2025-04-17.RData",
             INPUT.DIR = "<FULL_FILENAME FILE LOCATION>",
             OUTPUT.DIR = "<DESIRED ROUNDED FILE LOCATION>")

## End(Not run)
```

`runjagsdata.to.list` *Convert a runjags-class object to a list*

Description

Convert a runjags-class object to a list

Usage

```
runjagsdata.to.list(runjagsdata.in)
```

Arguments

`runjagsdata.in` (`runjags` Object) MCMC results from `autorun.jags`.

Value

A list object containing MCMC results from the provided runjags object.

`sample_verification` *Add Sample Verification Column (Level-2)*

Description

This function takes in a level-1 data frame and an exclusion list and returns a level-2 data frame with a verification column. The verification column contains either "Y", indicating the row is good for analysis, or messages contained in the exclusion list for why the data rows are excluded. If an exclusion list is not provided, all rows are assumed to be good for use in further analyses and are verified with "Y".

Usage

```
sample_verification(
  FILENAME,
  data.in,
  exclusion.info,
  assay,
  output.res = FALSE,
  INPUT.DIR = NULL,
  OUTPUT.DIR = NULL,
  verbose = TRUE
)
```

Arguments

FILENAME	(Character) A string used to identify the output level-1 file. "<FILENAME>-<assay>-Level1.tsv".
data.in	(Data Frame) A level-1 data frame from the format functions.
exclusion.info	(Data Frame) A data frame containing the variables and values of the corresponding variables to exclude rows. See details for full explanation.
assay	(Character) A string indicating what assay data the input file is. Valid input is one of the following: "Clint", "fup-UC", "fup-RED", or "Caco-2". This argument only needs to be specified when importing input data set with FILENAME or exporting a data file.
output.res	(Logical) When set to TRUE, the resulting data frame (level-2) will be exported to the user's per-session temporary directory or OUTPUT.DIR (if specified) as a .tsv file. (Defaults to FALSE.)
INPUT.DIR	(Character) Path to the directory where the input level-1 file exists. If NULL, looking for the input level-1 file in the current working directory. (Defaults to NULL.)
OUTPUT.DIR	(Character) Path to the directory to save the output file. If NULL, the output file will be saved to the user's per-session temporary directory or INPUT.DIR if specified. (Defaults to NULL.)
verbose	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

The 'exclusion.info' should be a data frame with the following columns:

Variables	level-1 variable(s) used to filter rows for exclusion
Values	Value(s) to exclude
Message	Simple explanation for the exclusion

When filtering on multiple variable-value pairs, the character input for "Variables" and "Values" should be separated by a vertical bar "|" , and the variable-value pairs should match. See demonstration in Examples, Scenario 1.

NOTE: Currently if NA's exist in a variable of interest for 'verification' assignments, then that variable cannot be used for assigning verification. Thus, either alternative variable-value pairs will need to be used in lieu of variable with missing values, or (though less ideal) "manual coding" adjustments in the verification column may be necessary.

If the output level-2 data frame is chosen to be exported and an output directory is not specified, it will be exported to the user's R session temporary directory. This temporary directory is a per-session directory whose path can be found with the following code: `tempdir()`. For more details, see https://www.collinberke.com/blog/posts/2023-10-24-til-temp-directories/.

As a best practice, `INPUT.DIR` (when importing a .tsv file) and/or `OUTPUT.DIR` should be specified to simplify the process of importing and exporting files. This practice ensures that the exported files can easily be found and will not be exported to a temporary directory.

Value

A level-2 data frame with a verification column.

Author(s)

Zhihui (Grace) Zhao

Examples

```
level1 <- invitroTKstats::clint_L1

# Scenario 1: Pass in data.in and exclusion.info data frame from R session

# Create a exclusion criteria data frame
# Use the excluded samples found in \code{invitroTKstats::clint_L2_heldout}
# If more than one variable is used to define a set of samples to be excluded,
# enter them as one string, separate the Variables with a vertical bar, "|",
# and do the same for Values.

excluded_level2 <- invitroTKstats::clint_L2_heldout

exclusion_criteria <- data.frame(
  Variables = paste("Compound.Name", "Lab.Sample.Name", sep = "|"),
 Values = paste(excluded_level2[,"Compound.Name"], excluded_level2[,"Lab.Sample.Name"], sep = "|"),
  Message = excluded_level2[,"Verified"]
  )

# Run the verification function.
my.level2 <- sample_verification(data.in=level1,
                                 exclusion.info = exclusion_criteria,
                                 output.res = FALSE)

# Scenario 2: Import 'tsv' as input data and do not pass in an exclusion.info data frame

## Not run:
# Write the level-1 file to some folder
# Will need to replace <desired level-1 FOLDER> with desired export folder location.
# The <desired level-1 FOLDER> needs to already exist.
```

```
write.table(level1,
file=here::here("<desired level-1 FOLDER>/Smeltz-Clint-Level1.tsv"),
sep="\t",
row.names=FALSE,
quote=FALSE)

# Run the verification function.
# Specify the path to import level-1 data with INPUT.DIR.
# Will need to replace INPUT.DIR = <desired level-1 FOLDER> with chosen output
# folder location from above
# If no exclusion.info data frame is used, will label all samples as verified.
# A level-2 file is also exported to INPUT.DIR when OUTPUT.DIR is not specified.
my.level2 <- sample_verification(FILENAME="Smeltz",
assay="Clint", INPUT.DIR = here::here("<desired level-1 FOLDER>"))

## End(Not run)
```

`scientific_10` *Formatting function for X-axis in log10-scale*

Description

Formatting function for X-axis in log10-scale

Usage

```
scientific_10(x)
```

Arguments

`x`	(Character) String to be formatted.

Value

Text with desired expression. Replace any scientific e notation to ten notation, simplify 10^01 to 10 and 10^0 to 1.

std.catcols *Standard Data Catalog (Data Guide) Columns*

Description

Standardized column names for data catalogs (i.e. data guides) used for collecting the minimum information to merge level-0 data files.

Usage

```
std.catcols
```

Format

A named character vector containing the default/standard column names for data catalogs, where the element names are the corresponding ‘create_catalog‘ arguments.

summarize_table *Creates a Summary Table of Mass-Spectrometry (MS) Data*

Description

This function creates and returns a list containing summary counts from the provided data frame containing mass-spectrometry (MS) data, MS calibration, chemical identifiers, and measurement type. The list includes the number of observations, unique chemicals, unique measurements in the input data table, and a vector of chemicals that have repeated observations. If a vector of data types is specified in the argument `req.types`, the function also checks if each chemical has observations for every measurement type included in the vector for each chemical-calibration pair. If it does, the chemical is said to have a complete data set. Otherwise, it has an incomplete data set. The number of complete and incomplete datasets, for each chemical, are returned in the output list. The input data frame can be level-1 (or level-2) Caco-2 data, ultracentrifugation (UC) data, rapid equilibrium dialysis (RED) data, or hepatocyte clearance (Clint) data. See the Details section for measurement type and annotation tables used in each assay.

Usage

```
summarize_table(
  input.table,
  dtxsid.col = "DTXSID",
  compound.col = "Compound.Name",
  cal.col = "Calibration",
  type.col = "Sample.Type",
  req.types = NULL,
  verbose = TRUE
)
```

Arguments

`input.table`	(Data Frame) A data frame (level-1 or level-2) containing mass-spectrometry peak areas, indication of chemical identity, and measurement type. The data frame should contain columns with names specified by the following arguments:
`dtxsid.col`	(Character) Column name of `input.table` containing EPA's DSSTox Structure ID (`http://comptox.epa.gov/dashboard`). (Defaults to "DTXSID".)
`compound.col`	(Character) Column name of `input.table` containing the test compound. (Defaults to "Compound.Name".)
`cal.col`	(Character) Column name of `input.table` containing the MS calibration. Calibration typically uses indices or dates to represent if the analyses were done on different machines on the same day or on different days with the same MS analyzer. (Defaults to "Calibration".)
`type.col`	(Character) Column name of `input.table` containing the sample type (see tables in Details). (Defaults to "Sample.Type".)
`req.types`	(Character Vector) A vector of character strings containing measurement types. If a vector is specified, each chemical-calibration pair will be checked if it has observations for all of the measurement types in the vector. (Defaults to `NULL`.)
`verbose`	(*logical*) Indicate whether printed statements should be shown. (Default is TRUE.)

Details

Sample types used in ultracentrifugation (UC) data collected for calculation of chemical fraction unbound in plasma (Fup) should be annotated as follows:

Calibration Curve	CC
Ultracentrifugation Aqueous Fraction	AF
Whole Plasma T1h Sample	T1
Whole Plasma T5h Sample	T5

Samples types used in rapid equilibrium dialysis (RED) data collected for calculation of chemical fraction unbound in plasma (Fup) should be annotated as follows:

No Plasma Blank (no chemical, no plasma)	NoPlasma.Blank
Plamsa Blank (no chemical, just plasma)	Plasma.Blank
Plasma well concentration	Plasma
Phosphate-buffered well concentration	PBS
Time zero plasma concentration	T0
Plasma stability sample	Stability
Acceptor compartment of the equilibrium evaluation	EC_acceptor
Donor compartment of the equilibrium evaluation (chemical spiked side)	EC_donor
Calibration Curve	CC

Sample types in hepatocyte clearance (Clint) data should be annotated as follows:

Blank	Blank
Hepatocyte incubation concentration	Cvst
Inactivated Hepatocytes	Inactive
Calibration Curve	CC

Samples types used in Caco-2 data to calculate membrane permeability should be annotated as follows:

Blank with no chemical added	Blank
Target concentration added to donor compartment at time 0 (C0)	D0
Donor compartment at end of experiment	D2
Receiver compartment at end of experiment	R2

Value

A list containing the summary counts from the input data table. The list includes the number of observations, the number of unique chemicals, the number of unique measurements, the number of chemicals with complete data sets, the number of chemicals with incomplete data sets, and the number of chemicals with repeated observations.

Author(s)

John Wambaugh

Examples

```
library(invitroTKstats)
# Smeltz et al. (2020) data:
##  Clint ##
summarize_table(
  input.table = invitroTKstats::clint_L2,
  req.types = c("Blank", "Cvst")
  )
## Fup RED ##
summarize_table(
  input.table = invitroTKstats::fup_red_L2,
  req.types=  c("Plasma", "PBS", "Plasma.Blank", "NoPlasma.Blank")
  )
## Fup UC ##
summarize_table(
  input.table = invitroTKstats::fup_uc_L2,
  req.types = c("CC", "T1", "T5", "AF")
  )
# Honda et al. () data:
## Caco2 ##
summarize_table(
  input.table = invitroTKstats::caco2_L2,
  req.types=c("Blank","D0","D2","R2")
  )
```

Index

∗ **datasets**
caco2_cheminfo, 8
caco2_L0, 9
caco2_L1, 10
caco2_L2, 11
caco2_L3, 12
clint_cheminfo, 36
clint_L0, 36
clint_L1, 38
clint_L2, 39
clint_L2_heldout, 40
clint_L3, 42
clint_L4, 43
clint_PREJAGS, 44
fup_red_cheminfo, 78
fup_red_L0, 79
fup_red_L2, 82
fup_red_L2_heldout, 83
fup_red_L3, 84
fup_red_L4, 85
fup_red_PREJAGS, 86
fup_uc_PREJAGS, 95
L1.common.col, 98
std.catcols, 112
∗ **data**
fup_red_L1, 80
fup_uc_cheminfo, 88
fup_uc_L0, 88
fup_uc_L1, 89
fup_uc_L2, 91
fup_uc_L2_heldout, 92
fup_uc_L3, 93
fup_uc_L4, 94
.check_all_miss_cols, 4
.check_char_cols, 4
.check_no_miss_cols, 5
.check_num_cols, 5
.check_std_colnames_in_data, 6

build_mydata_clint, 6
build_mydata_fup_red, 7
build_mydata_fup_uc, 7

caco2_cheminfo, 8
caco2_L0, 9
caco2_L1, 10
caco2_L2, 11
caco2_L3, 12
calc_caco2_point, 14, *107*
calc_clint, 17
calc_clint_point, 20
calc_fup_red, 23
calc_fup_red_point, 26
calc_fup_uc, 29
calc_fup_uc_point, 32
check_catalog, 35
clint_cheminfo, 36
clint_L0, 36
clint_L1, 38
clint_L2, 39
clint_L2_heldout, 40
clint_L3, 42
clint_L4, 43
clint_PREJAGS, 44
create_catalog, 45
create_chem_table, 47
create_method_table, 49

extract_level1_fup_uc, 51

format_caco2, *15*, 53
format_clint, *18*, *21*, 60, *104*
format_fup_red, *24*, *27*, 66
format_fup_uc, *30*, *33*, 72, *105*
fup_red_cheminfo, 78
fup_red_L0, 79
fup_red_L1, 80
fup_red_L2, 82
fup_red_L2_heldout, 83
fup_red_L3, 84

fup_red_L4, 85
fup_red_PREJAGS, 86
fup_uc_cheminfo, 88
fup_uc_L0, 88
fup_uc_L1, 89
fup_uc_L2, 91
fup_uc_L2_heldout, 92
fup_uc_L3, 93
fup_uc_L4, 94
fup_uc_PREJAGS, 95

Heaviside, 96

initfunction_clint, 96
initfunction_fup_red, 97
initfunction_fup_uc, 97

L1.common.col, 98
L1.common.cols *(*L1.common.col*)*, 98
lm, *19*, *21*

merge_level0, 98

plot_clint, 103
plot_fup_uc, 104

round_output, 105
runjagsdata.to.list, 108

sample_verification, 108
scientific_10, 111
std.catcols, 112
summarize_table, 112