

R Functions for NIST List of Possible PFAS

2/8/2023

Function Description

For the acquisition and useful application of the suspect list, additional functions were generated for R to process the dataset.

Set up R environment To open Microsoft Excel files, you must install the 'xlsx' package.

```
install.packages("xlsx")
```

```
source("fn/get_suspectlist.R")
get_suspectlist("PFAS Suspect List.xlsx")
# You can designate the list be saved as a different file
# name if you want.
```

get_suspectlist: To download the most recent version of the PFAS suspect list, use this function.

```
source("fn/open_suspectlist.R")
suspectlist <- open_suspectlist("PFAS Suspect List.xlsx")
## use the file location if this is different
```

open_suspectlist: Open the suspect list into the R environment (you must do this before any other steps)

```
source("fn/add_ionization.R")
suspectlist_w_ionization <- add_ionization(suspectlist, ionstate = c("M-H",
  "M-H2O+H", "M+H", "M+K", "M+Na"))
# These are all the possible ion states
write.csv(suspectlist_w_ionization, "suspectlist_w_ionization.csv",
  row.names = FALSE)
```

add_ionization: calculates the exact m/z value for indicated ionization states.

```
source("fn/calculate_residual.R")
exactmasses <- readRDS("fn/exactmasses.RDS")
suspectlist_w_residuals <- add_residual(suspectlist, rep_unit = "CF2",
  exactmasses, adduct = "neutral")
# adduct can be 'neutral', 'M-H', 'M-H2O+H', 'M+H', 'M+K',
# 'M+Na', 'M+', or 'M-'
write.csv(suspectlist_w_residuals, "suspectlist_w_residuals.csv",
  row.names = FALSE)
```

add_residual: calculates the residual mass after removing the repeating units (default = CF2)

```
source("fn/expand_additional.R")
suspectlist_w_additional <- expand_additional(suspectlist)
write.csv(suspectlist_w_additional, "suspectlist_w_additional.csv",
          row.names = FALSE)
```

expand_additional: all delimited information stored in the **ADDITIONAL** column is expanded as additional columns. **NOTE**: There will be a significant number of new columns with NA values for most compounds.

```
source("fn/get_properties.R")
suspectlist_w_properties <- get_properties(suspectlist)
write.csv(suspectlist_w_properties, "suspectlist_w_properties.csv",
          row.names = FALSE)
```

get_properties: using RDKit (<https://www.rdkit.org/>), calculates multiple 2D descriptors from the suspect list compounds. Requires the installation of RDKit and the use of the *tidychem* R Package. Definitions of the descriptors can be found at: <https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>

NIST License Statement

This data/work was created by employees of the National Institute of Standards and Technology (NIST), an agency of the Federal Government. Pursuant to title 17 United States Code Section 105, works of NIST employees are not subject to copyright protection in the United States. This data/work may be subject to foreign copyright.

The data/work is provided by NIST as a public service and is expressly provided “AS IS.” NIST MAKES NO WARRANTY OF ANY KIND, EXPRESS, IMPLIED OR STATUTORY, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT AND DATA ACCURACY. NIST does not warrant or make any representations regarding the use of the data or the results thereof, including but not limited to the correctness, accuracy, reliability or usefulness of the data. NIST SHALL NOT BE LIABLE AND YOU HEREBY RELEASE NIST FROM LIABILITY FOR ANY INDIRECT, CONSEQUENTIAL, SPECIAL, OR INCIDENTAL DAMAGES (INCLUDING DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, LOSS OF BUSINESS INFORMATION, AND THE LIKE), WHETHER ARISING IN TORT, CONTRACT, OR OTHERWISE, ARISING FROM OR RELATING TO THE DATA (OR THE USE OF OR INABILITY TO USE THIS DATA), EVEN IF NIST HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

To the extent that NIST may hold copyright in countries other than the United States, you are hereby granted the non-exclusive irrevocable and unconditional right to print, publish, prepare derivative works and distribute the NIST data, in any medium, or authorize others to do so on your behalf, on a royalty-free basis throughout the world.

You may improve, modify, and create derivative works of the data or any portion of the data, and you may copy and distribute such modifications or works. Modified works should carry a notice stating that you changed the data and should note the date and nature of any such change. Please explicitly acknowledge the National Institute of Standards and Technology as the source of the data: Data citation recommendations are provided at <https://www.nist.gov/open/license>.

Permission to use this data is contingent upon your acceptance of the terms of this agreement and upon your providing appropriate acknowledgments of NIST’s creation of the data/work.