Electronic Supplementary Information for:

Exploring Open Cheminformatics Approaches for Categorizing Per- and Polyfluoroalkyl Substances (PFASs)

Bo Sha[1,§], Emma L. Schymanski[2,*,§], Christoph Ruttkies[3], Ian T. Cousins[1], Zhanyun Wang[4]*

[1]Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University, SE-10691, Stockholm, Sweden, ORCID: 0000-0002-2176-0709
[2] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg. ORCID: 0000-0001-6868-8145
[3] Department Biochemistry of Plant Interactions, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany. ORCID: 0000-0002-8621-8689
[4] Chair of Ecological Systems Design, Institute of Environmental Engineering, ETH Zürich, 8093 Zürich, Switzerland, ORCID: 0000-0001-9914-7659
[§] shared first authors.
* corresponding authors: ELS: emma.schymanski@uni.lu
ZW: zhanyun.wang@ifu.baug.ethz.ch

Contents

This file contains instructions on how to use the splitPFAS tool mentioned in the main article. The electronic supplementary information (ESI) also includes - in separate files:

- Tables containing the ClassyFire results
- MetFrag2.4.5-Tools.jar (the jar file associated with the splitPFAS method, which can be used to follow the instructions here)
- An R script detailing how the results presented in the paper were calculated.

Instructions: how to use splitPFAS

Java is required to use the SplitPFAS tool. More information (including how to build the jar file) could be found at https://github.com/ipb-halle/MetFragRelaunched/tree/master/MetFragTools/src/main/java/de/ipbhalle/metfrag/split

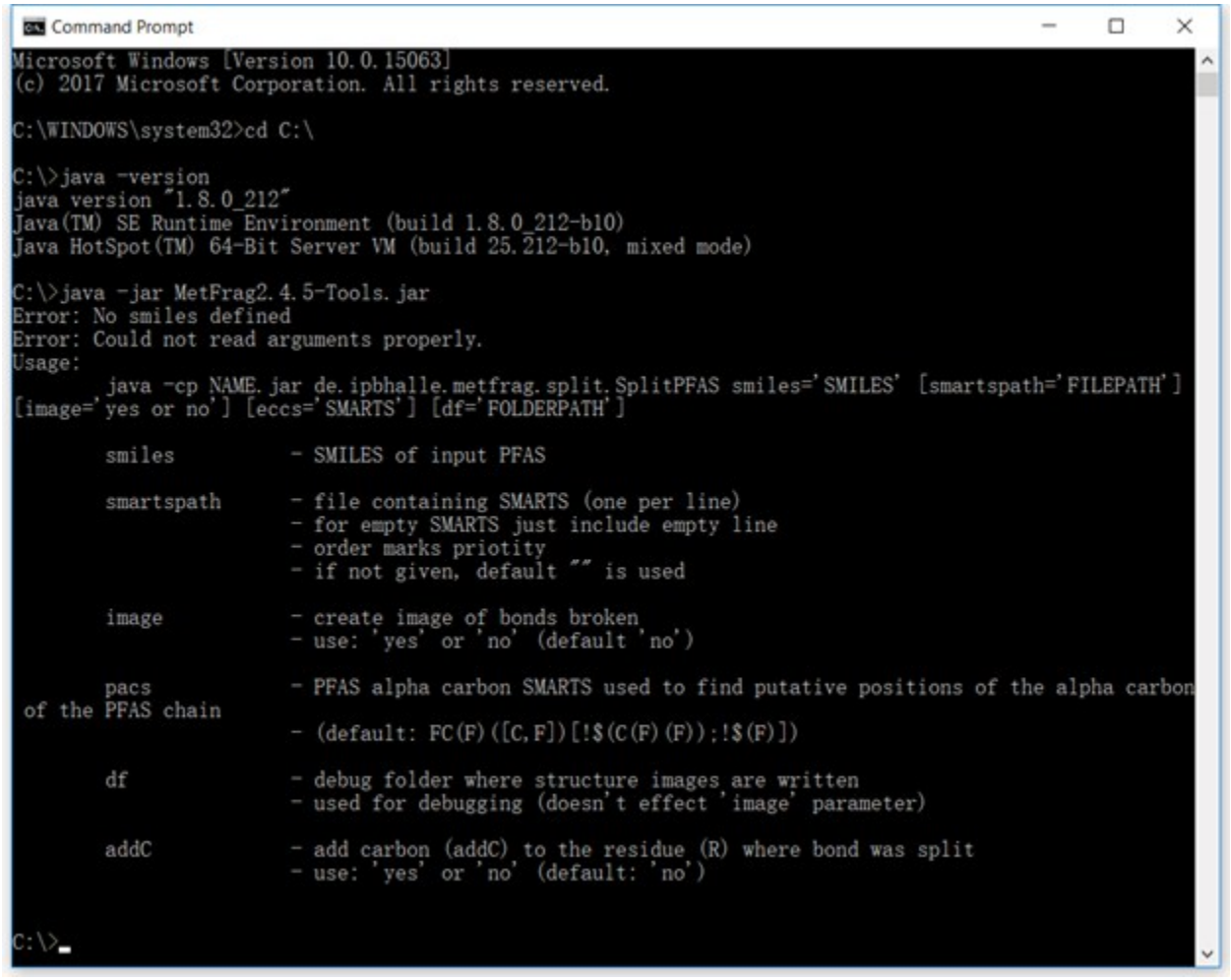
1. Usage

The splitPFAS tool can be run using commands such as the following either the Windows Command Prompt (first two examples) or in Linux/Mac (bottom), see screenshot below:

```
java -cp NAME.jar de.ipbhalle.metfrag.split.SplitPFAS smiles='SMILES'
[smartspath='FILEPATH'] [image='yes or no']
```

```
my_file_path>          java          -jar          MetFrag2.4.5-Tools.jar
smiles="OC(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F"
smartspath="splitPFAS_SMARTs.txt"
```

```
user@debian-user:~/temp$      java      -jar      MetFrag2.4.5-Tools.jar
smiles='OC(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F'
smartspath='splitPFAS_SMARTs.txt'
```



2. Parameters

NAME.jar – the name of the jar file, either downloaded from the ESI (name as shown above) or built according to the abovementioned link.

smiles – SMILES of input PFAS. The SMILES should be provided either within a pair of double quotes (windows), single quotes (Linux) - see above - or without quotes, e.g. smiles= OC(=O)C(F)(F)C(F)(F)C(F)(F)F or smiles="OC(=O)C(F)(F)C(F)(F)C(F)(F)F".

smartspath – Optional. The file containing SMARTS. Each SMARTS in the file should occupy one line. The order of the SMARTS indicates the priority. An example of the SMARTS file is shown in the screenshot below. If the file is not given, "" is used by default, which means the input SMILES would be split at the end of the perfluorinated carbon chain (C_nF_{2n+1}-). The file name and path should be provided either in double quote (windows), single quotes (Linux) or without quotes.

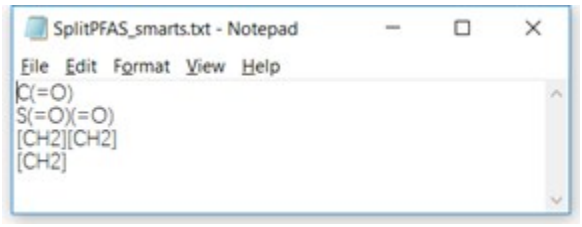


image – Optional. 'yes' or 'no' ('no' by default). If image="yes", a depict of the input molecule is created and the bond where the molecule split is highlighted.

pacs – Optional. pacs stands for PFAS Alpha Carbon SMARTS. This parameter is used to find putative positions of the alpha carbon of the PFAS chain (see manuscript for more details). Default: FC(F)([C,F])[!$(C(F)(F));!$(F)]).

df – Optional. debug folder where structure images are written (doesn't affect 'image' parameter).

addC – Optional. 'yes' or 'no' ('no' by default). Add a methyl group (C) to the residue (R) where the bond was split. Note: this function is still under development.

3. Output

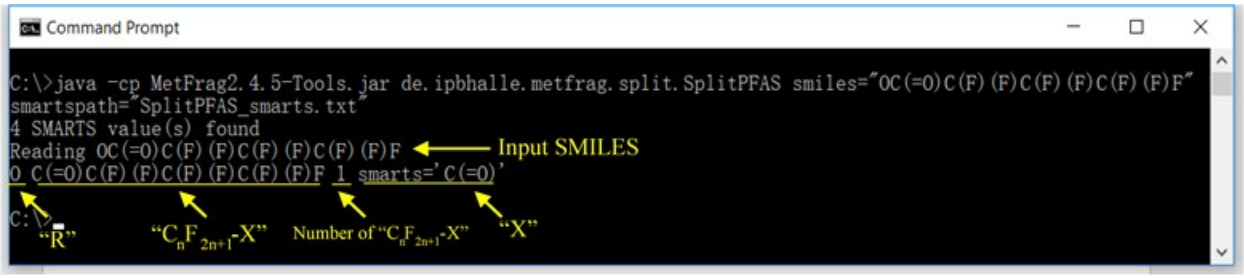
The output of the SplitPFAS tool contains three lines:

Line 1: indicating how many SMARTS is found in the SMARTS file.

Line 2: the input SMILES

Line 3: the splitting result, including "R", "C_nF_{2n+1}-X", number of "C_nF_{2n+1}-X" part and the splitting smarts ("X") separated by a space. If there are more than one "R" group and "C_nF_{2n+1}-X" part, they will be separated by "|".

An example screenshot is provided below. Input SMILES was "OC(=O)C(F)(F)C(F)(F)C(F)(F)F". Both the jar file (MetFrag2.4.5-Tools.jar) and the SMARTS file (splitPFAS_smarts.txt) were stored in the "C:\" folder.



4. Use SplitPFAS tool in R

Details could be found at https://github.com/schymane/RChemMass/blob/master/R/SplitPFAS.R