NAME

RDKitFilterPAINS.py - Filter PAINS molecules

SYNOPSIS

RDKitFilterPAINS.py [--infileParams <Name,Value,...>] [--mode <filter or count>] [--outfileParams <Name,Value,...>] [--painsMode <All, A, B or C>] [--negate <yes or no>] [--overwrite] [-w <dir>] -i <infile> -o <outfile>

RDKitFilterPAINS.py -h | --help | -e | --examples

DESCRIPTION

Filter Pan-assay Interference molecules (PAINS) [Ref 130 - 131] from an input file by performing a substructure search using SMARTS pattern specified in MAYACHEMTOOLS/lib/data/PAINSFilter.csv file and write out appropriate molecules to an output file or simply count the number of filtered molecules.

The supported input file formats are: SD (.sdf, .sd), SMILES (.smi, .csv, .tsv, .txt)

The supported output file formats are: SD (.sdf, .sd), SMILES (.smi)

OPTIONS

-e, --examples

Print examples.

-h, --help

Print this help message.

-i, --infile <infile>

Input file name.

--infileParams <Name,Value,...> [default: auto]

A comma delimited list of parameter name and value pairs for reading molecules from files. The supported parameter names for different file formats, along with their default values, are shown below:

```
SD: removeHydrogens,yes,sanitize,yes,strictParsing,yes
SMILES: smilesColumn,1,smilesNameColumn,2,smilesDelimiter,space,
    smilesTitleLine,auto,sanitize,yes
```

Possible values for smilesDelimiter: space, comma or tab.

-m, --mode <filter or count> [default: filter]

Specify whether to filter the matched molecules and write out the rest of the molecules to an outfile or simply count the number of matched molecules marked for filtering.

-n, --negate <yes or no> [default: no]

Specify whether to filter molecules not matching the PAINS filters specified by SMARTS patterns.

-o, --outfile <outfile>

Output file name.

--outfileParams <Name,Value,...> [default: auto]

A comma delimited list of parameter name and value pairs for writing molecules to files. The supported parameter names for different file formats, along with their default values, are shown below:

```
SD: compute2DCoords,auto,kekulize,no
SMILES: kekulize,no,smilesDelimiter,space, smilesIsomeric,yes,
    smilesTitleLine,yes
```

Default value for compute2DCoords: yes for SMILES input file; no for all other file types.

--overwrite

Overwrite existing files.

-p, --painsMode <All, A, B, or C> [default: All]

Specify PAINS filter family type to used for filtering molecules.

-w, --workingdir <dir>

Location of working directory which defaults to the current directory.

EXAMPLES

To count the number of molecules not containing any substructure corresponding to PAINS SMARTS patterns and write out a SMILES file, type:

```
% RDKitFilterPAINS.py -i Sample.smi -o SampleOut.smi
```

To only count the number of molecules not containing any substructure corresponding to PAINS SMARTS patterns without writing out any file file, type:

```
% RDKitFilterPAINS.py -m count -i Sample.sdf -o SampleOut.smi
```

To count the number of molecules containing any substructure corresponding to PAINS SMARTS patterns and write out a SD file with computed 2D coordinates, type:

```
% RDKitFilterPAINS.py -n yes -i Sample.smi -o SampleOut.sdf
```

To count the number of molecules not containing any substructure corresponding to PAINS SMARTS patterns family of Type A in a CSV SMILS file and write out a SD file, type:

```
% RDKitFilterPAINS.py --painsMode A --infileParams
  "smilesDelimiter,comma,smilesTitleLine,yes,smilesColumn,1,
  smilesNameColumn,2" --outfileParams "compute2DCoords,yes"
  -i SampleSMILES.csv -o SampleOut.sdf
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitConvertFileFormat.py, RDKitSearchSMARTS.py

COPYRIGHT

The functionality available in this script is implemented using RDKit, an open source toolkit for cheminformatics developed by Greg Landrum.

This file is part of MayaChemTools.

MayaChemTools is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.