NAME

RDKitEnumerateCompoundLibrary.py - Enumerate a virtual compound library

SYNOPSIS

RDKitEnumerateCompoundLibrary.py [--compute2DCoords <yes or no>] [--infileParams <Name,Value,...>] [--mode <RxnByName or RxnBySMIRKS>] [--outfileParams <Name,Value,...>] [--overwrite] [--prodMolNames <UseReactants or Sequential>] [--rxnName <text>] [--rxnNamesFile <FileName or auto>] [--smirksRxn <text>] [--sanitize <yes or no>] [-w <dir>] -i <ReactantFile1,...> -o <outfile>

RDKitEnumerateCompoundLibrary.py [--rxnNamesFile <FileName or auto>] -l | --list

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

DESCRIPTION

Perform a combinatorial enumeration of a virtual library of molecules for a reaction specified using a reaction name or SMIRKS pattern and reactant input files.

The SMIRKS patterns for supported reactions names [Ref 134] are retrieved from file, ReactionNamesAndSMIRKS.csv, available in MayaChemTools data directory. The current list of supported reaction names is shown below:

'1,2,4_triazole_acetohydrazide', '1,2,4_triazole_carboxylic_acid_ester', 3_nitrile_pyridine, Benzimidazole_derivatives_aldehyde, Benzimidazole_derivatives_carboxylic_acid_ester, Benzofuran, Benzothiazole, Benzothiophene, Benzoxazole_aromatic_aldehyde, Benzoxazole_carboxylic_acid, Buchwald_Hartwig, Decarboxylative_coupling, Fischer_indole, Friedlaender_chinoline, Grignard_alcohol, Grignard_carbonyl, Heck_non_terminal_vinyl, Heck_terminal_vinyl, Heteroaromatic_nuc_sub, Huisgen_Cu_catalyzed_1,4_subst, Huisgen_disubst_alkyne, Huisgen_Ru_catalyzed_1,5_subst, Imidazole, Indole, Mitsunobu_imide, Mitsunobu_phenole, Mitsunobu_sulfonamide, Mitsunobu_tetrazole_1, Mitsunobu_tetrazole_2, Mitsunobu_tetrazole_3, Mitsunobu_tetrazole_4, N_arylation_heterocycles, Negishi, Niementowski_quinazoline, Nucl_sub_aromatic_ortho_nitro, Nucl_sub_aromatic_para_nitro, Oxadiazole, Paal_Knorr_pyrrole, Phthalazinone, Pictet_Spengler, Piperidine_indole, Pyrazole, Reductive_amination, Schotten_Baumann_amide, Sonogashira, Spiro_chromanone, Stille, Sulfon_amide, Suzuki, Tetrazole_connect_regioisomer_1, Tetrazole_connect_regioisomer_2, Tetrazole_terminal, Thiazole, Thiourea, Triaryl_imidazole, Urea, Williamson_ether, Wittig

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

-c, --compute2DCoords <yes or no> [default: yes]

Compute 2D coordinates of product molecules before writing them out.

-i, --infiles <ReactantFile1, ReactantFile2...>

Comma delimited list of reactant file names for enumerating a compound library using reaction SMIRKS. The number of reactant files must match number of reaction components in reaction SMIRKS. All reactant input files must have the same format.

--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, MOL: 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. These parameters apply to all reactant input files, which must have the same file format.

-e, --examples

Print examples.

-h, --help

Print this help message.

-l, --list

List available reaction names along with corresponding SMIRKS patterns without performing any enumeration.

-m, --mode <RxnByName or RxnBySMIRKS> [default: RxnByName]

Indicate whether a reaction is specified by a reaction name or a SMIRKS pattern. Possible values: RxnByName or RxnBySMIRKS.

-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: kekulize,no
SMILES: kekulize,no,smilesDelimiter,space, smilesIsomeric,yes,
    smilesTitleLine,yes
```

-p, --prodMolNames <UseReactants or Sequential> [default: UseReactants]

> Generate names of product molecules using reactant names or assign names in a sequential order. Possible values: UseReactants or Sequential. Format of molecule names: UseReactants - <ReactName1>_<ReactName2>..._Prod<Num>; Sequential - Prod<Num>

--overwrite

> Overwrite existing files.

-r, --rxnName <text>

> Name of a reaction to use for enumerating a compound library. This option is only used during 'RxnByName' value of '-m, --mode' option.

--rxnNamesFile <FileName or auto> [default: auto]

> Specify a file name containing data for names of reactions and SMIRKS patterns or use default file, ReactionNamesAndSMIRKS.csv, available in MayaChemTools data directory.

> Reactions SMIRKS file format: RxnName,RxnSMIRKS.

> The format of data in local reaction names file must match format of the reaction SMIRKS file available in MayaChemTools data directory.

-s, --smirksRxn <text>

> SMIRKS pattern of a reaction to use for enumerating a compound library. This option is only used during 'RxnBySMIRKS' value of '-m, --mode' option.

--sanitize <yes or no> [default: yes]

> Sanitize product molecules before writing them out.

-w, --workingdir <dir>

> Location of working directory which defaults to the current directory.

EXAMPLES

To list all available reaction names along with their SMIRKS pattern, type:

```
% RDKitEnumerateCompoundLibrary.py -l
```

To perform a combinatorial enumeration of a virtual compound library corresponding to named amide reaction, Schotten_Baumann_amide and write out a SMILES file type:

```
% RDKitEnumerateCompoundLibrary.py -r Schotten_Baumann_amide
  -i 'SampleAcids.smi,SampleAmines.smi' -o SampleOutCmpdLibrary.smi
```

To perform a combinatorial enumeration of a virtual compound library corresponding to an amide reaction specified using a SMIRKS pattern and write out a SD file containing sanitized molecules, computed 2D coordinates, and generation of molecule names from reactant names, type:

```
% RDKitEnumerateCompoundLibrary.py -m RxnBySMIRKS
  -s '[O:2]=[C:1][OH].[N:3]>>[O:2]=[C:1][N:3]'
  -i 'SampleAcids.smi,SampleAmines.smi' -o SampleOutCmpdLibrary.sdf
```

To perform a combinatorial enumeration of a virtual compound library corresponding to an amide reaction specified using a SMIRKS pattern and write out a SD file containing unsanitized molecules, without generating 2D coordinates, and a sequential generation of molecule names, type:

```
% RDKitEnumerateCompoundLibrary.py -m RxnBySMIRKS -c no -s no
  -p Sequential -s '[O:2]=[C:1][OH].[N:3]>>[O:2]=[C:1][N:3]'
  -i 'SampleAcids.smi,SampleAmines.smi' -o SampleOutCmpdLibrary.sdf
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitConvertFileFormat.py, RDKitFilterPAINS.py, RDKitSearchFunctionalGroups.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.