

NAME

RDKitCalculateMolecularDescriptors.py - Calculate 2D/3D molecular descriptors

SYNOPSIS

```
RDKitCalculateMolecularDescriptors.py [--autocorr2DExclude <yes or no>] [--fragmentCount <yes or no>] [
--descriptorNames <Name1,Name2,...>] [--infileParams <Name,Value,...>] [--mode <2D, 3D, All...>] [--outfileParams
<Name,Value,...>] [--overwrite] [--precision <number>] [--smilesOut <yes or no>] [-w <dir>] -i <infile> -o <outfile>
```

```
RDKitCalculateMolecularDescriptors.py -l | --list
```

```
RDKitCalculateMolecularDescriptors.py -h | --help | -e | --examples
```

DESCRIPTION

Calculate 2D/3D molecular descriptors for molecules and write them out to a SD or CSV/TSV text file.

The complete list of currently available molecular descriptors may be obtained by using '-l, --list' option. The names of valid 2D, fragment count, and 3D molecular descriptors are shown below:

2D descriptors: Autocorr2D, BalabanJ, BertzCT, Chi0, Chi1, Chi0n - Chi4n, Chi0v - Chi4v, EState_VSA1 - EState_VSA11, ExactMolWt, FpDensityMorgan1, FpDensityMorgan2, FpDensityMorgan3, FractionCSP3, HallKierAlpha, HeavyAtomCount, HeavyAtomMolWt, Ipc, Kappa1 - Kappa3, LabuteASA, MaxAbsEStateIndex, MaxAbsPartialCharge, MaxEStateIndex, MaxPartialCharge, MinAbsEStateIndex, MinAbsPartialCharge, MinEStateIndex, MinPartialCharge, MolLogP, MolMR, MolWt, NHOHCount, NOCount, NumAliphaticCarbocycles, NumAliphaticHeterocycles, NumAliphaticRings, NumAromaticCarbocycles, NumAromaticHeterocycles, NumAromaticRings, NumHAcceptors, NumHDonors, NumHeteroatoms, NumRadicalElectrons, NumRotatableBonds, NumSaturatedCarbocycles, NumSaturatedHeterocycles, NumSaturatedRings, NumValenceElectrons, PEOE_VSA1 - PEOE_VSA14, RingCount, SMR_VSA1 - SMR_VSA10, SlogP_VSA1 - SlogP_VSA12, TPSA, VSA_EState1 - VSA_EState10, qed

FragmentCount 2D descriptors: fr_Al_COO, fr_Al_OH, fr_Al_OH_noTert, fr_ArN, fr_Ar_COO, fr_Ar_N, fr_Ar_NH, fr_Ar_OH, fr_COO, fr_COO2, fr_C_O, fr_C_O_noCOO, fr_C_S, fr_HOCCN, fr_Imine, fr_NH0, fr_NH1, fr_NH2, fr_N_O, fr_Ndealkylation1, fr_Ndealkylation2, fr_Nhpyrrole, fr_SH, fr_aldehyde, fr_alkyl_carbamate, fr_alkyl_halide, fr_allylic_oxid, fr_amide, fr_amidine, fr_aniline, fr_aryl_methyl, fr_azide, fr_azo, fr_barbitur, fr_benzene, fr_benzodiazepine, fr_bicyclic, fr_diazo, fr_dihydropyridine, fr_epoxide, fr_ester, fr_ether, fr_furan, fr_guanido, fr_halogen, fr_hdrzine, fr_hdrzone, fr_imidazole, fr_imide, fr_isocyan, fr_isothiocyan, fr_ketone, fr_ketone_Topliss, fr_lactam, fr_lactone, fr_methoxy, fr_morpholine, fr_nitrile, fr_nitro, fr_nitro_arom, fr_nitro_arom_nonortho, fr_nitroso, fr_oxazole, fr_oxime, fr_para_hydroxylation, fr_phenol, fr_phenol_noOrthoHbond, fr_phos_acid, fr_phos_ester, fr_piperidine, fr_piperzine, fr_priamide, fr_prisulfonamd, fr_pyridine, fr_quatN, fr_sulfide, fr_sulfonamd, fr_sulfone, fr_term_acetylene, fr_tetrazole, fr_thiazole, fr_thiocyan, fr_thiophene, fr_unbrch_alkane, fr_urea

3D descriptors: Asphericity, Autocorr3D, Eccentricity, GETAWAY, InertialShapeFactor, MORSE, NPR1, NPR2, PMI1, PMI2, PMI3, RDF, RadiusOfGyration, SphericityIndex, WHIM

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

The supported output file formats are: SD File (.sdf, .sd), CSV/TSV (.csv, .tsv, .txt)

OPTIONS

-a, --autocorr2DExclude <yes or no> [default: yes]

Exclude Autocorr2D descriptor from the calculation of 2D descriptors.

-f, --fragmentCount <yes or no> [default: yes]

Include 2D fragment count descriptors during the calculation. These descriptors are counted using SMARTS patterns specified in FragmentDescriptors.csv file distributed with RDKit. This option is only used during '2D' or 'All' value of '-m, --mode' option.

-d, --descriptorNames <Name1,Name2,...> [default: none]

A comma delimited list of supported molecular descriptor names to calculate. This option is only used during 'Specify' value of '-m, --mode' option.

-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, 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.

-l, --list

List molecular descriptors without performing any calculations.

-m, --mode <2D, 3D, All, FragmentCountOnly, or Specify> [default: 2D]

Type of molecular descriptors to calculate. Possible values: 2D, 3D, All or Specify. The name of molecular descriptors must be specified using '-d, --descriptorNames' for 'Specify'. 2D descriptors also include 1D descriptors. The structure of molecules must contain 3D coordinates for the calculation of 3D descriptors.

-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

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

-p, --precision <number> [default: 3]

Floating point precision for writing the calculated descriptor values.

-s, --smilesOut <yes or no> [default: no]

Write out SMILES string to CSV/TSV text output file.

--overwrite

Overwrite existing files.

-w, --workingdir <dir>

Location of working directory which defaults to the current directory.

EXAMPLES

To compute all available 2D descriptors except Autocorr2D descriptor and write out a CSV file, type:

```
% RDKitCalculateMolecularDescriptors.py -i Sample.smi -o SampleOut.csv
```

To compute all available 2D descriptors including Autocorr2D descriptor and excluding fragment count descriptors, and write out a TSV file, type:

```
% RDKitCalculateMolecularDescriptors.py -m 2D -a no -f no  
-i Sample.smi -o SampleOut.tsv
```

To compute all available 3D descriptors and write out a SD file, type:

```
% RDKitCalculateMolecularDescriptors.py -m 3D -i Sample3D.sdf  
-o Sample3DOut.sdf
```

To compute only fragment count 2D descriptors and write out a SD file file, type:

```
% RDKitCalculateMolecularDescriptors.py -m FragmentCountOnly  
-i Sample.sdf -o SampleOut.sdf
```

To compute all available 2D and 3D descriptors including fragment count and Autocorr2D and write out a CSV file, type:

```
% RDKitCalculateMolecularDescriptors.py -m All -a no -i Sample.sdf  
-o SampleOut.csv
```

To compute a specific set of 2D and 3D descriptors and write out a write out a TSV file, type:

```
% RDKitCalculateMolecularDescriptors.py -m specify  
-d 'MolWt,MolLogP,NHOHCount, NOCount,RadiusOfGyration'  
-i Sample3D.sdf -o SampleOut.csv
```

To compute all available 2D descriptors except Autocorr2D descriptor for molecules in a CSV SMILES file, SMILES strings in column 1, name in column 2, and write out a SD file without calculation of 2D coordinates, type:

```
% RDKitCalculateMolecularDescriptors.py --infileParams  
"smilesDelimiter,comma,smilesTitleLine,yes,smilesColumn,1,  
smilesNameColumn,2" --outfileParams "compute2DCoords,no"  
-i SampleSMILES.csv -o SampleOut.sdf
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitCalculateRMSD.py, RDKitCompareMoleculeShapes.py, RDKitConvertFileFormat.py, RDKitGenerateConformers.py,
RDKitPerformMinimization.py

COPYRIGHT

Copyright (C) 2018 Manish Sud. All rights reserved.

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.