

NAME

RDKitCalculateRMSD.py - Calculate RMSD between molecules

SYNOPSIS

```
RDKitCalculateRMSD.py [--calcRMSD <RMSD, BestRMSD>] [--infileParams <Name,Value,...>] [--maxI ters <number>] [
--mode <OneToOne, AllToAll, FirstToAll>] [--overwrite] [-w <dir>] -r <reffile> -p <probefile> -o <outfile>
```

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

DESCRIPTION

Calculate Root Mean Square Distance (RMSD) between a set of similar molecules in reference and probe input files. The RDKit function fails to calculate RMSD values for dissimilar molecules. Consequently, a text string 'None' is written out as a RMSD value for dissimilar molecule pairs.

The supported input file formats are: Mol (.mol), SD (.sdf, .sd)

The supported output file formats are: CSV (.csv), TSV (.tsv, .txt)

OPTIONS

-c, --calcRMSD <RMSD, BestRMSD> [default: RMSD]

Methodology for calculating RMSD values. Possible values: RMSD, BestRMSD. During BestRMSMode mode, the RDKit 'function AllChem.GetBestRMS' is used to align and calculate RMSD. This function calculates optimal RMSD for aligning two molecules, taking symmetry into account. Otherwise, the RMSD value is calculated using 'AllChem.AlignMol function' without changing the atom order. A word to the wise from RDKit documentation: The AllChem.GetBestRMS function will attempt to align all permutations of matching atom orders in both molecules, for some molecules it will lead to 'combinatorial explosion'.

--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
```

--maxI ters <number> [default: 50]

Maximum number of iterations to perform for each molecule pair during minimization of RMSD values. This option is ignored during BestRMSD mode.

-m, --mode <OneToOne, AllToAll, FirstToAll> [default: OneToOne]

Specify how molecules are handled in reference and probe input files during calculation of RMSD between reference and probe molecules. Possible values: OneToOne, AllToAll and AllToFirst. For OneToOne mode, the number of molecules in reference file must be equal to the number of molecules in probe file. The RMSD is calculated for each pair of molecules in the reference and probe file and written to the output file. For AllToAll mode, the RMSD is calculated for each reference molecule against all probe molecules. For FirstToAll mode, however, the RMSD is only calculated for the first reference molecule against all probe molecules.

-e, --examples

Print examples.

-h, --help

Print this help message.

-p, --probefile <probefile>

Probe input file name.

-r, --reffile <reffile>

Reference input file name.

-o, --outfile <outfile>

Output file name for writing out RMSD values. Supported text file extensions: csv or tsv.

--overwrite

Overwrite existing files.

-w, --workingdir <dir>

Location of working directory which defaults to the current directory.

EXAMPLES

To calculate RMSD between pair of molecules in reference and probe molecules in 3D SD files and write out a CSV file containing calculated RMSD values along with appropriate molecule IDs, type:

```
% RDKitCalculateRMSD.py -r Sample3DRef.sdf -p Sample3DProb.sdf
-o SampleOut.csv
```

To calculate RMSD between all molecules in reference and probe molecules in 3D SD files and write out a CSV file containing

calculated RMSD values along with appropriate molecule IDs, type:

```
% RDKitCalculateRMSD.py -m AllToAll -r Sample3DRef.sdf -p
Sample3DProb.sdf -o SampleOut.csv
```

To calculate best RMSD between first molecule in reference all probe molecules in 3D SD files and write out a TSV file containing calculated RMSD values along with appropriate molecule IDs, type:

```
% RDKitCalculateRMSD.py -m FirstToAll --calcRMSD BestRMSD -r
Sample3DRef.sdf -p Sample3DProb.sdf -o SampleOut.tsv
```

To calculate RMSD between all molecules in reference and probe molecules in 3D SD files without removing hydrogens and write out a TSV file containing calculated RMSD values along with appropriate molecule IDs, type:

```
% RDKitCalculateRMSD.py -m AllToAll --infileParams
"removeHydrogens,no" -r Sample3DRef.sdf -p Sample3DProb.sdf
-o SampleOut.tsv
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitCalculateMolecularDescriptors.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.