NAME

        RDKitCompareMoleculeShapes.py - Compare shapes of molecules

SYNOPSIS

        RDKitCompareMoleculeShapes.py [--alignment <Open3A, CrippenOpen3A>] [--distance <Tanimoto, Protrude, Both>] [--infileParams <Name,Value,...>] [--maxIters <number>] [--mode <OneToOne, AllToAll, FirstToAll>] [--overwrite] [-w <dir>] -r <reffile> -p <probefile> -o <outfile>

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

DESCRIPTION

        Compare shapes of molecules between a set of molecules in reference and probe input files. The molecules are aligned using either Open 3DAlign or Crippen Open 3DAlign before calculating shape Tanimoto and protrude distances.

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

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

OPTIONS

        -a, --alignment <Open3A, CrippenOpen3A> [default: Open3A]

            Alignment methodology to use for aligning molecules before calculating Tanimoto and protrude shape distances. Possible values: Open3A or CrippenOpen3A. Open 3DAlign (Open3A) [Ref 132] overlays molecules based on MMFF atom types and charges. Crippen Open 3DAlign (CrippenOpen3A) uses Crippen logP contributions to overlay molecules.

        -d, --distance <Tanimoto, Protrude, Both> [default: Both]

            Shape comparison distance to calculate for comparing shapes of molecules. Possible values: Tanimoto, Protrude, or Both. Shape Tanimoto distance takes the volume overlay into account during the calculation of distance. Shape protrude distance, however, focuses on the volume mismatch.

        --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

        --maxIters <number> [default: 50]

            Maximum number of iterations to perform for each molecule pair during alignment.

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

            Specify how molecules are handled in reference and probe input files during comparison of shapes between reference and probe molecules. Possible values: OneToOne, AllToAll and AllToFirst. For OneToOne mode, the molecule shapes are calculated for each pair of molecules in the reference and probe file and the shape distances are written to the output file. For AllToAll mode, the shape distances are calculated for each reference molecule against all probe molecules. For FirstToAll mode, however, the shape distances are 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 shape distances. 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 perform shape alignment using Open3A methodology between pair of molecules in reference and probe input in 3D SD files, calculate both Tanimoto and protrude distances, and write out a CSV file containing calculated distance values along with appropriate molecule IDs, type:

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

To perform shape alignment using Crippen Open3A methodology between all molecules in reference and probe molecules in 3D SD files, calculate only Tanimoto distance, and write out a TSV file containing calculated distance value along with appropriate molecule IDs, type:

```
% RDKitCompareMoleculeShapes.py  -m AllToAll -a CrippenOpen3A -d Tanimoto
  -r Sample3DRef.sdf -p Sample3DProb.sdf -o SampleOut.csv
```

To perform shape alignment using Open3A methodology between first reference molecule against all probe molecules in 3D SD files without removing hydrogens , calculate both Tanimoto and protrude distances, and write out a CSV file containing calculated distance values along with appropriate molecule IDs, type:

```
% RDKitCompareMoleculeShapes.py -m FirstToAll -a Open3A -d Both
  --infileParams "removeHydrogens,no" -r Sample3DRef.sdf
  -p Sample3DProb.sdf -o SampleOut.csv
```

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

RDKitCalculateRMSD.py, RDKitCalculateMolecularDescriptors.py, RDKitConvertFileFormat.py, RDKitGenerateConformers.py, RDKitPerformMinimization.py

COPYRIGHT