

NAME

RDKitAlignMolecules.py - Align molecules by RMSD or shape

SYNOPSIS

```
RDKitAlignMolecules.py [--alignment <Open3A, CrippenOpen3A, RMSD, BestRMSD>] [--infileParams
<Name,Value,...>] [--maxI ters <number>] [--mode <OneToOne, FirstToAll>] [ --outfileParams
<Name,Value,...> ] [--overwrite] [-w <dir>] -r <reffile> -p <probefile> -o <outfile>
```

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

DESCRIPTION

Perform alignment between a set of similar molecules in reference and probe input files. The molecules are aligned either by Root Mean Square Distance (RMSD) between molecules or overlying their shapes (Open3A or CrippenOpen3A). The RDKit function fails to calculate RMSD values for dissimilar molecules. Consequently, unaligned probe molecules are written to the output file for dissimilar molecule pairs.

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

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

OPTIONS

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

Alignment methodology to use for aligning molecules. Possible values: Open3A, CrippenOpen3A, RMSD, BestRMSD.

The Open3A and CrippenOpen3A allow alignment of molecules using their shapes 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.

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, FirstToAll> [default: OneToOne]

Specify how molecules are handled in reference and probe input files during alignment of molecules between reference and probe molecules. Possible values: OneToOne and FirstToAll. For OneToOne mode, the alignment is performed for each pair of molecules in the reference and probe file, and the aligned probe molecule is written the output file. For FirstToAll mode, the alignment is only performed between the first reference molecule against all probe molecules.

-e, --examples

Print examples.

-h, --help

Print this help message.

-o, --outfile <outfile>

Output file name for writing out aligned probe molecules values. Supported file extensions: sdf or sd.

--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,yes

-p, --probefile <probefile>

Probe input file name.

-r, --reffile <reffile>

Reference input file name.

--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 pairs of molecules in reference and probe input 3D SD files and write out a SD file containing aligned molecules, type:

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

To perform alignment using RMSD methodology between pairs of molecules in reference and probe input 3D SD files and write out a SD file containing aligned molecules, type:

```
% RDKitAlignMolecules.py -a RMSD -r Sample3DRef.sdf -p Sample3DProb.sdf  
-o SampleOut.sdf
```

To perform alignment using Open3A methodology between first reference molecule against all probe molecules in 3D SD files without removing hydrogens , and write out a SD file containing aligned molecules, type:

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

AUTHOR

Manish Sud(msud@san.rr.com)

SEE ALSO

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

COPYRIGHT

Copyright (C) 2022 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.