Vfilter v1.0

CONTENTS

OVERVIEW

Vfilter analyzes multiple conformations of a molecule and removes conformations that are identical to within user-specified tolerances. Both global and local chemical symmetries are accounted for when comparing a pair of conformations. For example, if two conformations differ only by a 180º rotation of a benzene ring (see Figure), the local symmetry of the benzene is detected and applied so that the two conformations are recognized as identical. Alternate resonance forms, but not tautomers, are accounted for when identifying symmetries.

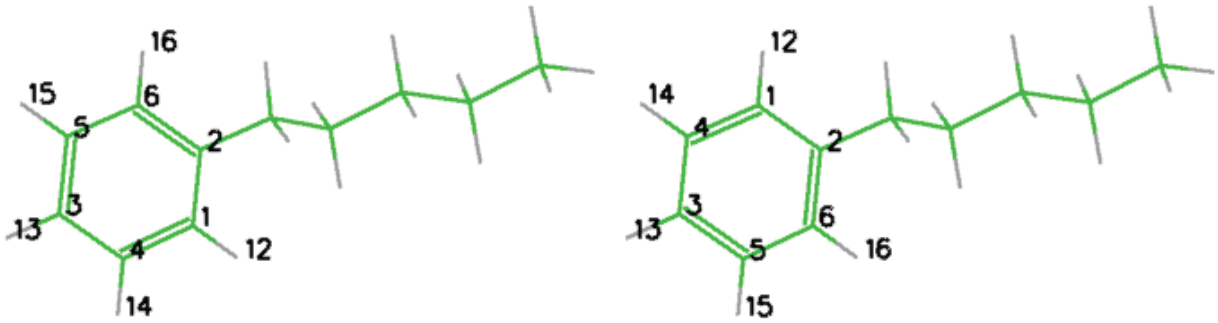

Figure: Two conformations that are identical except for a chemically irrelevant benzene flip.

The inputs are:

- an SDfile containing multiple conformations of the same molecule. (See www.mdli.com/downloads/literature/ctfile.pdf for a description of the SDfile format.)
- command-line options

The outputs are:

- a log file containing a summary of the run information.
- an SDfile containing the filtered conformations.

COMMAND-LINE EXECUTION

Vfilter is invoked via the following command line:

```
vfilter.exe <source(.sdf)> [-e cuttoffE] [-et energyTol] [-dt maxDist] [-at
maxAng] [-ts setupTimeLimit] [-rl limitRes] [-m maxConfs] [-log logFile]
[-out filteredFile]
```

(Items in square brackets are optional.)

Options:

`-e` Energy cutoff

`cutoffE` Conformations with energies more than `cutoffE` kcal/mol above the energy of the most stable conformation will not be filtered or written out.

Default: Filter all conformations; that is, `cutoffE` is infinite.

`-et` Energy tolerance

`energyTol` Molecules that differ in energy by `energyTol` kcal/mol or more are automatically considered to be distinct. Applying this criterion speeds filtering by reducing the number of conformations that need to be compared geometrically. If there is no <energy> block in the sdf input, this option should not be used.

Default: 1.0 kcal/mol

Conformations that pass the energy filter are then filtered based on geometry. First, if the difference between the radii of gyration of two molecules is larger than `maxDist`, or any corresponding angles or dihedrals differ more than `maxAng`, the two conformations are regarded as different. If the two conformations are still considered the same at this point, then their atom-by–atom, symmetry-corrected root-mean-square deviation (RMSD) is computed. If this RMSD is greater than `maxDist Å`, then the higher energy conformation is eliminated.

`-dt` Distance tolerance

`maxDist` Two conformations whose radii of gyration differ by more than `maxDist` Å, are considered different and not subjected to further comparisons. Also, two conformations are considered different if their symmetry-corrected, atom-by-atom RMSD is greater than or equal to `maxDist` Å.

Default: 0.2 Angstroms.

`-at` Angle/Dihedral tolerance

`maxAng` If any pair of corresponding bond angles or dihedral angles of two molecules differ by more than `maxAng`, the two conformations are regarded as different and no further comparisons are made. Symmetry is accounted for in determining which angles correspond to each other.

Default: 30 degrees

`-ts` Molecule setup time limit

`setupTimeLimit` Time limit for molecule setup, in seconds, where setup includes and calculation of partial charges (used in symmetry calculation). Occcasionally a molecule will exceed this limit, usually due to an especially complex resonant system. A greater time limit may be used or the depth of the resonance search limited with (-rl). Another alternative, if it is acceptable, is to use the ignore symmetry option (-ns).

Default: 120 seconds

`-rl` Resonance generation limit

`limitRes` Stop making new resonance forms if any generation has more than specified number of states. Otherwise resonance will run until the setup time limit is reached.

Default: 1000 generations

`-m` Maximum number of output conformations

`maxConfs` The maximum number of unique conformations of lowest energy will be written to the output file. This option is only meaningful it there is an <energy> block in the sdf input containing the energy of the conformation.

Default: No limit on number of output conformations

`-log` Log filename

`logFile` Name for the log file containing the filtering information.

Default: `<input sdf/mol filename>_vfilter.log`

`-out` File of filtered conformations

`filteredFile` Name for the file containing the filtered conformations.

Default: `<input sdf/mol filename>_filtered.sdf`

Stopping a Run in Progress

A run in progress can be cleanly stopped by placing into the working directory an empty file with the name of the log file but with ".log" replaced by ".end". For example, if the log file is "diversity_vfilter.log", the run will be stopped by placing a file named "diversity_vfilter.end" in the same directory as the log file.

The Linux "touch" command provides an easy way to create this file; e.g., "`touch diversity_vfilter.end`".

The Files section of the log file includes a reminder of the name and location of the file required to stop the calculation; see the "Stop file:" line.

EXAMPLE

A sample input file, example.sdf, is included, along with output. The input file contains conformations generated with by but not filtered in **Vconf**.

The command:

```
vfilter.exe example.sdf
```

`generates the output files example_filtered.sdf` and `example_vfilter.log.`