

User's Manual

VCONF IDENTIFIES stable conformations of drug-like molecules, beginning from either 2D or 3D molecular representations in the MDL SD or MOL file format. It provides many options to help you obtain the results you want, but is still simple to use. A reasonable set of defaults is built in.

GETTING STARTED

To carry out a full conformational search, use the command `vconf.exe SDfilename|Molfilename`.

For example, to carry out the calculation on all molecules in the file "molecules.sdf":

```
vconf.exe molecules.sdf
```

Only for molecules one through ten:

```
vconf.exe molecules.sdf -l 10
```

For molecules 11 through 20:

```
vconf.exe molecules.sdf -f 11 -l 20
```

For each input molecule, the conformation with the most stable set of ring conformations is placed in `molecules_vconf.sdf`. In addition, Vconf creates a separate SDfile for each molecule with its full set of conformations. For a list of command-line options, type:

```
vconf.exe -help
```

You can monitor the progress of a calculation by viewing the log file `molecules.log` in a text editor.

A running calculation can be terminated cleanly, after the current molecule is complete, by creating an empty file whose name matches that of the log file, but with '.log' replaced with '.end'. For example, if the log file is "molecules.log" you can stop the run by creating the file "molecules.end" in the directory from which the calculation was started. In Linux, this can be done with the command "touch molecules.end". The log file contains the name and location of the file that must be created to stop the calculation. (See the "Stop file" line in the "Files" section of the log file.)

EXAMPLE FILES

This software distribution includes a sample input file, `example.sdf`, with two molecules, and corresponding output files. One of the molecules in the input file is processed successfully by Vconf with the default parameters. In order to illustrate error handling, the other molecule is designed to fail with default parameters because the formal charges are not specified in the input file. The command

```
vconf.exe example.sdf
```

generates the outputs:

```
example.log, example_vconf.sdf, XN901_confs.sdf
```

The second molecule in the input file can be successfully processed by using the `-fc` option:

```
vconf.exe example.sdf -f 2 -l 2 -fc -log examplefc.log -out examplefc_vconf.sdf
```

This generates the outputs:

```
examplefc.log, examplefc_vconf.sdf, acetylcholine_confs.sdf
```

The included shell script `test.sh` (`test.bat` for windows) runs these commands automatically. Two sets of example results are provided in the `example_results` subdirectory. One set was produced on a 2.6GHz Opteron linux system and the other, named with '_2', was generated on a 2.8GHz Pentium 4 Windows XP machine. Due to processor performance differences your results might be slightly different from those in this subdirectory. Your variation should be of the same order as that present in the example results.

OVERVIEW

Key features of Vconf include:

- Thorough search mode constructs initial 3D conformations with various stable ring conformations, and then uses these conformations as starting points for cycles of distortion and energy minimization with the powerful Tork algorithm.
- Rapid prep mode only constructs and outputs the initial 3D conformations used in the search mode; these are excellent starting points for docking and scoring calculations.
- User may designate atoms to be kept fixed during the search.
- Chirality and cis/trans isomers may be locked
- Sophisticated conformational filtering eliminates repeat and unwanted conformations.
- Extensive command-line options provide detailed control.

The main inputs are: a MOL or SD with one or more molecules; command-line options controlling the interpretation of the input SDfile, the search, and the conformational filter.

The main outputs are: an SDfile containing the best structure (lowest energy in search mode, lowest sum of ring energies in prep mode) found for each molecule in the input SDfile; for each molecule in the input SDfile, an output SDfile with all conformations found for the molecule; a log file containing sum-

mary information and error messages.

Vconf sets up each molecule by adding missing hydrogens, identifying alternate resonance forms, assigning force field atom-types, computing atomic partial charges, and identifying rings and ring systems. It then generates one or more stable conformations of each ring system, and then builds 3D conformations of the entire molecule by linking ring conformers and relaxing the resulting conformations. In prep mode, no additional conformation searching is done and Vconf moves directly to the filtering stage. In search mode, the initial 3D conformations are used as starting points for a whole-molecule, energy-based conformational search. Once the search is complete, the conformations are filtered to remove repeats with an algorithm that accounts for local and global molecular symmetries. Also, during the entire search procedure, any new conformation is rejected if it violates specified atom chiralities or double bond cis/trans isomer states.

Vconf averaged 20 seconds/molecule for a search run and 3.9 seconds/molecule for a prep run on 2000 drug-like molecules from the Maybridge screening database, on a 2.4 GHz AMD Opteron 280 processor.

INPUT FILES

The main input file is either a standard MDL SDfile with multiple molecules, or a Molfile with one molecule. (For details of these formats, see [ctfile.pdf](#)).

If the GAFF forcefield is selected than prmtop and mol2 files for each molecule must also be provided.

THE GAFF FORCEFIELD

Vconf can carry out calculations using the GAFF forcefield if provided with prmtop and mol2 files generated by the Antechamber program. Antechamber is available from <http://amber.scripps.edu/antechamber/antechamber.html> for unix/linux systems. A pair of these files must be provided for each molecule to be processed. The runAntechamber program provided with linux version of the Vconf suite provides a simple way to generate these files for molecules in a given SDFfile. See the **-ff** option description below as well as the runAntechamber documentation for details.

GENERALIZED BORN SOLVATION

An implementation of the Hawkins 96 Generalized Born implicit solvation model is available. (*J. Phys. Chem.* 100:19824-19839, 1996) See the **-gb** option description below.

STEREOCHEMISTRY

Vconf can limit its output to include only conformations with correct stereoisomers, based upon atom stereochemistry information in the input SDfile. This information can be read from the parity column of the atom block of the input SDfile, where 1 and 2 are alternate parities, and 0 and 3 are read as undefined. Atom chirality information can also be read from stereo bond information or atomic coordinates in the SDfile. Alternate stereo configurations of an atom are best explored by listing each variant as a separate molecule in the input file so that each stereoisomer is equally and fully sampled. An alternative approach

would be to turn off stereo checking with the **-sr n** or **-sr i** option, but Vconf is not designed to vary stereochemistry, so the sampling may not be adequate and this approach is therefore not recommended.

Vconf also can preserve the cis/trans stereoisomerism of double bonds; this information is read from either 2D or 3D coordinates in the input SDfile.

NITROGEN LONE PAIRS

It is sometimes important to prevent inversion of nitrogen centers during the conformational search. Vconf allows the chirality of a pyramidal nitrogen to be specified through addition of an explicit lone pair to the input molecule file. The chirality of the nitrogen can then be handled in the same way as the chirality of a carbon atom. To add an explicit lone pair:

1. Use the element "lp" to add the lone pair as a new row in the atom block.
2. Create an "<LP attachment>" data block as illustrated below.

```
> <LP attachment>
2 2
5 18 0
11 9 1
```

The two numbers on the first of the "LP attachment" line refer to the total number of lone pairs in the molecule and the total number of lone pair bonds. (Normally, the number of lone pairs equals the number of lone pair bonds.) Here "5 18 0" means that the lone pair listed as atom 18 is associated with a nitrogen listed as atom 5, and the pseudo-bond joining them is not associated with a stereo wedge, as indicated by the 0. Similarly, "11 9 1" means that the lone pair listed as atom 19 is associated with a nitrogen listed as atom 11 and that there is an "up" stereo bond from the nitrogen to the lone pair. (For a description of bond type designations see the MDL SDfile documentation. Note that Vconf will not include lone pairs in the atom block of the output SDfiles.

When computing atom parities, Vconf always considers a hydrogen atom to have a higher atom index than any non-hydrogen atom, in keeping with the SDfile specification. Vconf furthermore considers a lone pair to have an index higher than any non-hydrogen, but lower than any hydrogen. However, there is no standard treatment of lone pairs in the SDfile format, so it is recommended that the stereochemistry of chiral nitrogens be specified with stereo bonds, rather than with atom parities, in order to minimize ambiguity.

OUTPUT FILES

Vconf generates an SDfile designed to let the user quickly review the results (*SDfilename_vconf.sdf*), a set of files SDfiles, each containing all the conformations found for given molecule (the *moleculeName_confs.sdf* files), and a log file with additional information about the run.

Note: if more than one calculation is carried out in the same directory on the same molecules the moleculeName_

confs.sdf files will be overwritten. This will happen even if the user chooses different names for the log and `_vconf.sdf` files (see **-log** and **-out** options). It is therefore recommended that runs be done in separate directories.

SDfilename_vconf.sdf: The name of this file is based on that of the source SDfile. If the input file is “molecules.sdf”, then the summary output SDfile is named “molecules_vconf.sdf” by default. For each molecule, the output SDfile contains:

- for a search run: the lowest-energy conformation found in the search
- for a prep run: the structure formed from the lowest energy combination of ring conformations.

If the program is unable to process an input molecule, then “*SDfilename_vconf.sdf*” contains the conformation from the input file along with error messages describing the problem(s). (See the “Error and Warning Messages” section for further information.)

The “*SDfilename_vconf.sdf*” file also contains data blocks for each molecule with the energy of the conformation (<Energy>), the sum of the ring energies used to generate the conformation (<Ring Energy>), any error messages (<VC.Vconf.Error>) and the total CPU time spent on the molecule (<Time>).

moleculeName_confs.sdf: For each molecule in the input file, Vconf generates a separate SDfile, named “*moleculeName_confs.sdf*”, which contains all conformations generated for the molecule, along with the energy of the conformation and the sum of ring energies in the initial conformation that led to the conformation. Here ‘*moleculeName*’ is the molecule name found on line one of the header block of the molecule in the input SDfile. If this field is missing, then the SDfile for *i*-th molecule in the input SDfile is named “*SDfilename_mol_i_confs.sdf*”. If the molecule name consists of more than one word, then the first word is used to form the name of the SDfile. The conformations are sorted by conformational energy in a search run, and by the sum of ring energies in a prep run.

moleculeName_unfiltered.sdf: Vconf can write an additional file for each molecule with all unfiltered conformations; see **-u** option. This file follows the same conventions as the “*moleculeName_confs.sdf*” file. (Note that the “*moleculeName_unfiltered.sdf*” file does not include the mirror images automatically generated as part of the filtering process (see COMPUTATIONAL METHODOLOGY/Conformational Filtering) and therefore may contain fewer conformations than the “*moleculeName_confs.sdf*” file.)

SDfilename.log: This file records additional information about the run, including the version of the program, the locations and names of files used and generated by the program, the command line and values of the parameters used in the run, and information about each molecule processed, including:

- The molecule index and name. (See naming rules in the previous paragraph.)
- The parity assignment for each chiral atom in the molecule

- Error messages, if any
- The energy of the lowest energy conformation found
- The number of conformations found after filtering. (The number of unfiltered conformations is approximately equal to the number of searches plus the number initial 3D conformational builds.) (See Command Line Options for details.) Some searches, however do not yield conformations because they, for example, violate the input stereochemistry.
- Time used in processing the molecule
- Time used filtering conformations for the molecule

For “search” runs, additional information is included that is useful in determining if a more extensive run should be done on the molecule, including:

- How far into the search the lowest-energy conformation was found;
- How far into the search the last unique conformation was found;
- How far into the search 90% percent of the unique conformations were found.

For “prep” runs, additional information is included about the ring energies, including:

- Lowest sum of ring energies found for this molecule. (Note that the ring energies depend on what type of ring fragment was chosen for the run; see **-sub** option in Ring Options.)
- The total energy of the conformation with the lowest sum of ring energies. This is not necessarily the lowest energy conformation found. There are two reasons for this: 1) Since there is no attempt to find the lowest energy conformation in a prep run, when the molecule is assembled with the lowest energy ring conformations it may be in a higher energy conformation than conformations assembled from higher energy ring conformations. 2) The ring conformation that is lowest in energy without its full attachments may not be the lowest in the fully assembled molecule.

The file “*SDfilename.log*” is updated continuously and thus provides current information on the progress of a calculation. If the default log file name is used, a new calculation for the same input SDfile will overwrite any existing log file, so it is advisable to specify a new log file name by using the “**-o logFile**” option on the command line.

Temporary files are written to the “/tmp” directory. Please make sure that there is at least 20MB of space in “/tmp” so that the run will proceed smoothly.

COMMAND LINE OPTIONS

Vconf's default settings are designed to yield good generic prep and search results for databases of small- to medium-sized drug-like molecules. In addition, many command-line options are available so you can further tune the calculations for your particular requirements. For a summary of all command line options, use the command "vconf -h"

General Options

-f *firstMol*

-l *lastMol*

Molecule range options.

firstMol Index of the first molecule in the SDfile to be processed. Default: 1

lastMol Index of the last molecule in the SDfile to be processed. Default: last molecule in the file.

-ff *forcefield*

Select forcefield to be used.

Possible values for forcefield:

gaff Use the GAFF forcefield. 'prmtop' and 'mol2' files for each molecule (generated by Antechamber) must be present in the same directory as the source sdf/mol file.

dreiding Use the modified Dreiding forcefield.

Default: dreiding

-gb

Use Generalized Born solvation

Turns on the Hawkins 96 Generalized Born implicit solvation model. (*J. Phys. Chem.* 100:19824-19839, 1996)

-h

Print help message.

Search Options

-m *mode*

Type of calculation

Possible values for mode:

search computes ring conformations, links rings to form initial 3D molecular conformations, relaxes these conformations, and then carries out a full-molecule conformational search starting from each initial conformation. Parameters controlling the individual ring searches are set with Ring Options (below).

prep computes ring conformations, links rings to form initial 3D molecular conformations, and relaxes these conformations. Parameters controlling the individual ring searches are set with Ring Options (below). This option is useful for quick 2D to 3D conversions and generating starting conformations for other applications, such as ligand-receptor docking algorithms, that treat

rings as rigid.

Default: search

-ns *numSteps*

Number of search steps

numSteps integer number of search steps, where a step consists of a distortion of the molecule followed by energy minimization, to be carried out for each molecule in a search calculation. The number of search steps is the chief determinant of the thoroughness and duration of a calculation in search mode, but this parameter does not affect a calculation in prep mode.

Default: 100

-ra *indx1 ... indxN*

Restrain atoms

indx1...indxN a list of atom numbers (e.g., "3 4 27 29") to be locked to their conformation in the input SDfile. At least 3 atoms must be listed, and these atoms must form a single connected fragment.

-seed *randomNumberOption*

Random number options

Possible values for *randomNumberOption*:

ran The ran option directs Vconf to generate a new set of random number seeds, based upon the system time, for use in the stochastic part of the calculation. The values of the four random number seeds are written to the log file and can be used with the following option to exactly reproduce a calculation. The random number seeds affect the Monte Carlo calculations used to relax the initial full molecule builds, and also affect the stochastic selection of "driver" combinations during Tork search steps.

s1, s2, s3, s4 Explicitly providing Vconf with 4 integer random number seeds allows a prior calculation to be reproduced exactly. The seeds used in a run are written to the log file.

-sw *searchWidth*

Search width

searchWidth In search mode, the initial 3D conformations built by stitching rings into the full molecule and relaxing the resulting conformations are used as starting points for full-molecule Tork searches. Each initial conformation is subjected to a preliminary Tork search, and the resulting conformations from all the initial conformations are compared. The one with the lowest energy is used as the single initial conformation for additional Tork search cycles. The *searchWidth* parameter allows the

user to specify what fraction of the total numStep search steps are allocated to the preliminary Tork searches from the initial 3D builds. A larger value of *searchWidth* causes Vconf to put more time into the preliminary broad exploration starting from the initial 3D builds. A smaller value of *searchWidth* causes Vconf to put more time into in-depth follow-up of the lowest energy conformation found during the preliminary searches. The value of *searchWidth* must be between 0 and 1.0.

Default: 0.5

Molecule Options

-fc

Formal charge assumption

Specifying **-fc** causes Vconf to assign formal charges to atoms with the assumption that all hydrogen atoms are explicitly represented in the SDfile. For example, the formal charge of an ammonium will be determined by the number of hydrogen and non-hydrogen atoms to which it is bonded, along with the bond orders, irrespective of any formal charge that might be specified in the atom block of the SDfile.

Default: The default assumption is to assume that formal charges are correctly given in the atom block of the SDfile. Vconf will then add hydrogens as needed to generate the correct valences. Using the **-fc** option will cause valence checking errors and/or abnormal formal charges if the SDfile does not include all required hydrogens.

-snp

Sulfonamide nitrogen conformation

This option causes Vconf to treat sulfonamide nitrogen atoms as pyramidal.

Default: Sulfonamide nitrogens are treated as planar.

-cp priorityString

Chiral read priority

This option allows you to control what chirality information will be read from the input SDfile for use in conformational filtering. An SDfile can specify chirality with the parity field in the atom block, with stereo bond information in the bond block, and/or with the 3D coordinates of the atoms. There is no guarantee that these three chirality specifications will be mutually consistent. This is an intrinsic limitation of the SDfile format. The **-cp** flag allows the user to specify which of these specifications Vconf can use to obtain chirality information, and to assign priorities to them for use in case of inconsistency or missing information.

The following options are available:

- p Use the parity field in the atom block.
- b Use the stereo bond information in the bond block.
- c Use 3D coordinates.

Any combination of these options is allowed. If an option is not specified, chirality information will not be retrieved by that method.

The order in which the options are listed on the command line determines their priorities: Vconf will attempt to recover chirality information from the first listed option, and will then fall back to the second and third, if they are listed. If the chirality of a chiral center cannot be assigned by any of the specified options, then Vconf writes a warning message to the log file and does not filter its output according to the chirality of this center. Vconf can detect and preserve the configuration of diastereomeric molecules. This function occasionally results in messages concerning the assignment of chirality to atoms which are not themselves chiral, but whose configuration, in combination with the configurations of other atoms, determines the diastereomer configuration of the molecule as a whole.

Examples:

-cp c Use only the 3D coordinates to determine chirality. If there are no 3D coordinates, or the 3D coordinates are ambiguous, write a warning to the log file and do not filter conformations by chirality at this center.

-cp pbc If the parity field is set to 1 or 2, use it to assign chirality. Otherwise, try to use stereo bond information. If this also is not available, use the 3D coordinates. If there is no unambiguous chirality information at all, write a warning to the log file and do not filter conformations by chirality at this center.

Default: pbc

-sr stereochemistryRestriction

Stereochemistry restrictions

Each new conformation generated by Vconf in the course of the calculation is immediately checked for compliance with stereoisomer restrictions, and is discarded if it violates any operative restriction. The **-sr** option controls which types of stereochemistry are read from the input SDfile and applied as restrictions.

Possible values for *stereochemistryRestriction*:

- b Filtering based upon both chirality and cis/trans isomerism.
- n No chirality or cis/trans filtering.
- c Filtering based upon only chirality
- i Filter based upon only double bond cis/trans isomerism.

Default: b

Note: Turning off these filters (e.g., **-sr n**) may allow some alternate stereoisomers to be generated, but the coverage of the various isomers may not be good. If you wish to search across multiple stereo and cis/trans isomers, the results will be best if each configuration is listed separately in the input SDfile.

Note: cis/trans isomers are determined from the 2D or 3D coordinates in the input SDfile. When the resulting cis/trans specifications are incorrect, extremely high energy conformations (>1000 kcal/mol) can result, and will be reported in the log file. Rarely, incorrect cis/trans specifications lead to distorted output conformations whose energies are still less than the 1000 kcal/mol output cutoff. If such a case is suspected, then: 1) check the cis/trans specification in the input SDfile, and/or 2) allow conformations with altered cis/trans isomerism to be written out by using: **-sr c** or **-sr n**, as this may allow Vconf to fix the isomer state.

-dc dielectricCoef

Dielectric coefficient

dielectricCoef Coefficient for the distance dependent dielectric constant. (See Computational Methodology.)

Default: 4.0

-nlp

Nitrogen lone pair

A pyramidal nitrogen atom single-bonded to three different substituents can be viewed as pseudo-chiral. If the SDfile includes an explicit lone pair for such a nitrogen, Vconf treats it as chiral according to the same rules used for chiral carbons. If the SDfile does not include an explicit lone pair for such a nitrogen atom, then the default is not to treat it as chiral. However, specifying **-nlp** on the command line causes Vconf to attempt to establish its chirality using the current chiral priorities (see **-cp** above), as detailed below. *There is considerable potential for ambiguity and inconsistency when automatically adding lone pairs. Vconf handles only relatively well-defined cases that have been found to yield generally good results for a number of different compound catalogs and databases.*

Atom parity: If the SDfile includes the parity of the nitrogen atom and the nitrogen has three different explicit substituents, the parity is interpretable because Vconf assumes the lone pair to have an index higher than any non-hydrogen atom but lower than any hydrogen atom. If only two substituents are explicit and the third is a hydrogen atom that Vconf must add, then the atom parity could still be interpreted, but tests on compound databases indicate that the results are unreliable. Vconf therefore treats such nitrogen atoms as having an unspecified chirality.

Stereo bonds: Vconf interprets stereo bond information for a chiral nitrogen atom only if the input conformation is 2D on the XY plane, the nitrogen has explicit bonds to three listed atoms, and one of the bonds is a stereo (wedge) bond. In this case, Vconf assumes a lone pair with the same XY coordinates as the stereo-bonded atom, but with an oppositely oriented wedge bond, and will then determine if the nitrogen parity is well-defined.

3D coordinates: Vconf uses the coordinates of the nitrogen's three explicit substituents to ascertain the orientation of the pyramidal nitrogen. The lone pair is then positioned to

establish a tetrahedral configuration, or one as near to tetrahedral as possible. If the nitrogen is essentially planar, as defined by an improper dihedral of less than 5 degrees, then the nitrogen will not be considered as "chiral" and its stereochemistry will not be locked.

Default: Do not interpret chirality information for pseudo-chiral nitrogen atoms unless the input file includes explicit lone pairs.

-rl resLimit

Resonance generation limit

resLimit Stop generating resonance forms if any generation has more than **resLimit** states (see COMPUTATIONAL METHODOLOGY/Molecule Setup). Otherwise resonance will run until the setup time limit is reached.

Default: 1000

Ring Options

As detailed in the Methodology section, Vconf begins processing a molecule by identifying all rings and ring systems ("ring fragments") with fewer atoms than a user-specified threshold, and constructing multiple stable conformations of each ring fragment. Combinations of the ring conformations are then stitched together with the rest of the molecule and the molecule is relaxed by torsional Monte Carlo followed by energy minimization. In prep mode, the resulting full molecule conformations are then filtered and written out. In search mode, the resulting conformations are then subjected to full-molecule conformational search. The Ring Options control the construction, conformational search, and output of ring conformations.

Note: Ring conformations cannot be generated for the rare molecule with more than 20 ring systems, and Vconf also will not generate conformations for a ring fragment with more than 200 atoms. For both of these cases, ring conformations can still be generated during the full-molecule search in search mode. Vconf also will not generate and use more than 40 conformations for a given ring fragment. Should the molecule consist en-

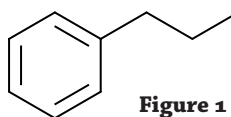


Figure 1

tirely of a single ring fragment (e.g. cyclohexane), then the ring templating procedure is not used, and only the full-molecule conformational search is conducted. In this case, the ring options listed below are ignored.

-sub subLevel

First level substituents

The atoms bonded to the ring atoms ("first level substituents") can strongly affect the conformational preferences of the ring. Vconf provides several different options for these important atoms. (Refer to the example molecule in **Figure**

1.)

Possible values for subLevel:

- h** Use only hydrogens as first level substituents. For the example molecule, the ring fragment would simply be cyclohexane.
- u** Use united atoms for first level substituents. For the example, the first level substituents would all be hydrogen atoms, except that atom 1 would be linked to a united-atom representation of a methyl group. This option provides accuracy at low computational cost.
- a** Use all-atom first level substituents. All-atom methylcyclohexane would be used as the ring fragment.

Default: u

-minrs *minRingSearchSteps*

Minimum number of search steps per ring fragment.

minRingSearchSteps minimum number of steps of distortion and minimization to be carried out for each ring fragment.

Default for “-m prep”: 50

Default for “-m search”: 1/2 of *numSteps* (see **-ns** and **-sw**)**-addrs** *addRingSearchSteps*

Maximum number of additional search cycles for each ring fragment

addRingSearchSteps if *minRingSearchSteps* of distortion and minimization have been carried out for a ring fragment without identifying any “good” conformations, additional sets of 50 steps will be carried out until at least one good conformation is found or until *addRingSearchSteps* step are completed. A ring conformation is considered good if its energy is less than 1000 kcal/mol and it satisfies all chirality and cis/trans criteria. (See Conformational Filtering Options.)

Default: 200

-maxrc *maxRingConfs*

Maximum number of ring conformations generated per fragment

maxRingConfs maximum number of ring conformation for each ring fragment. The value must be between 1 and 40.

Default: 5

-maxra *maxRingAtoms*

Maximum number of ring atoms

maxRingAtoms Independent ring conformations will only be generated for ring fragments that have *maxRingAtoms* or fewer at-

oms, excluding substituents. The value of *maxRingAtoms* must be in the range 3 to 200. The conformations of larger ring systems are explored during the full molecule conformational search in search mode.

The option to change the maximum size of ring fragment is useful in a number of situations in which search mode is used; e.g.:

- The conformation of a large flexible ring which forms part of a larger molecule may be searched more thoroughly by not generating conformations for the ring fragment on its own but instead allowing it to be searched during the full-molecule Tork search. Reducing *maxRingAtoms* below the size of the ring accomplishes this.
- When a molecule includes a mixture of large and small rings, the best results often are obtained if *maxRingAtoms* is set so that the smaller rings are searched independently of the full molecule while the larger rings are searched in the context of the entire molecule.
- If more than 40 conformations are desired for a ring system, then setting *maxRingAtoms* below the size of the ring system will prevent the ring system from being processed as a ring fragment and will thus allow its conformations to be thoroughly explored during the full molecule search.

Default: The lesser of 200 and the number of heavy atoms in the molecule

-re *cutoffRingE*

Ring energy cutoff

cutoffRingE Vconf discards ring conformations more than *cutoffRingE* kcal/mol above the energy of the most stable conformation.

Default: 5.0 kcal/mol

-klr

Keep lowest energy ring combination

Include the full molecule conformation with the lowest sum of ring energies in the subsequent Tork search even if this conformation has a higher total energy than other conformations (see Computational Methodology).

Note: This option is relevant only in search mode.

Conformational Filtering Options

Vconf filters out conformational repeats, and the filter accounts for symmetries when comparing conformations; for example, two conformations that differ only by a benzene flip are considered identical and only one is written to the output. Note that stereoisomer checking is done during the conformational search and is not affected by these filtering options. (Vfilter, VeraChem’s standalone program for conformational filtering, can be used to filter the output of any program for generating con-

formations. Vrms is another standalone utility for computing the differences between multiple conformations of the same molecule, while accounting for resonance and symmetry.)

-e *cutoffE*

Energy cutoff

cutoffE Conformations with energies more than *cutoffE* kcal/mol above the energy of the will not be written out.

Default for “-m prep”: 5 kcal/mol

Default for “-m search”: 50 kcal/mol

-et *energyTol*

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.

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 not different by these criteria, 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 of the two is eliminated.

-dt *maxDist*

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 Å.

-at *maxAng*

Angle/Dihedral tolerance

maxAng If any pair of corresponding bond angles or dihedral angles of two molecules differ by more than *maxAng* degrees, the two conformations are regarded as different and no further comparisons are made.

Default: 30.0 degrees.

-nf

Do not filter output.

Default: Filter output.

Note: Another option, **-u**, (see Output Options) causes Vconf to write an additional file with the unfiltered conformations which can then be filtered separately with Vfilter. This alter-

native is convenient if one wishes to experiment with different filtering options without rerunning Vconf.

Time Limits

Vconf provides for time limits on the initial setup of each molecule and on the conformational filtering of each molecule. However, there is no time limit for the conformational search itself. The time spent on the search is determined by the complexity of the molecule and the number of search steps requested.

-ts *setupTimeLimit*

Molecule setup time limit

setupTimeLimit Time limit for molecule setup, in seconds, where setup includes identification of ring fragments and alternate resonance forms, calculation of partial charges, and atom-typing. The occasional molecule that exceeds this time limit, usually due to an especially complex resonance system, is skipped to avoid delaying the run, and can be revisited later with a greater time limit or no time limit (see **-nt** option below).

Default: 120 seconds

-tf *filterTimeLimit*

Filtering time limit.

filterTimeLimit Time limit for filtering the conformations for a molecule. Conformations of the occasional molecule that exceeds the filtering limit, usually due to complex symmetries, are written to the output unfiltered. These conformations can then be filtered separately with the standalone program Vfilter.

Default: The default is for Vconf to set *filterTimeLimit* to *setupTimeLimit*

-nt

No time limits

No time limit is imposed on molecule setup or conformational filtering.

Default: Apply time limits.

Output Options

-u

Keep unfiltered conformations

Save an additional file of unfiltered conformations. For each molecule in the input SDfile, Vconf will write a separate file containing the unfiltered conformations and named “*moleculeName_unfiltered.sdf*”. (Note that the ‘_unfiltered.sdf’ file does not include the mirror images automatically generated as part of the filtering process (see COMPUTATIONAL METHODOLOGY/Conformational Filtering) and therefore may contain fewer conformations than the ‘_confs.sdf’ file.)

Default: Do not save unfiltered conformations.

-log *logFilename*

Log file

logFile Name of the log file (See Output Files).

Default: *SDfilename.log*

-out *lowestSdFilename*

File of lowest energy conformations

lowestSDF Name of output file with the lowest energy conformation of each molecule, in search mode, or the conformation with the lowest sum of ring energies, in prep mode.

Default: *SDFilename_vconf.sdf*

SUPPORTED ELEMENTS AND BOND TYPES

Hydrogen
Carbon
Oxygen
Nitrogen
Fluorine
Phosphorous
Sulfur
Chlorine
Bromine
Iodine

Single, double and triple bonds orders are supported in Vconf. Aromatic systems must be specified via single and double bonds, rather than by “aromatic” bonds.

ERROR AND WARNING MESSAGES

Error and warning messages for each molecule are written to the log file and to data blocks in *SDfile_vconf.sdf*. The following messages may be encountered:

1. Unable to parse molecule

The input SDfile or MOLfile cannot be interpreted for a given molecule. This usually results from problems with the file contents, such as missing or corrupt lines.

2. File format error

The input file for a given molecule appears to contain a violation of the SDfile or MOLfile format. This error is usually caused by numbers out of range. For example, the stereo parity column of the atom block should contain only 0, 1, 2, or 3; any other integer violates the file format.

3. Molecule has no atoms

The input file for the given molecule appears to contain no atom records.

4. Bond order not specified

This version of Vconf only supports input in which all bonds have an explicit integer bond order. Input files with bonds designated as “aromatic” or with an ambiguous order will generate this message.

5. Unsupported element

This message is generated when a molecule contains an element that is not currently supported by Vconf. (See Supported Elements and Bond Types.)

6. Unsupported atom type (valence error)

This message is generated when an atom cannot be matched to an atom type supported by Vconf due to what appears to be an incorrect valence.

7. Unsupported atom type (radical)

Radicals are not supported by this version of Vconf.

8. Unsupported atom type(s) (charger)

This message is generated when an atom is of a type not supported by Vconf due to lack of parameters for the charge generation algorithm.

9. Restrained atoms must all be in a single connected fragment

This message is generated when a section of the molecule is to be kept fixed during the search (**-ra**), but the specified atoms are not part of a single bonded cluster (i.e. each fixed atom must be connected to all others via bonds which pass only through other fixed atoms.)

Note: Error messages 10-14 address situations in which Vconf cannot find any low-energy conformations that satisfy the stereochemical requirements for chirality and/or cis/trans isomers, as specified in the input file. These error messages include suggestions for resolving the error. In general, this situation usually results from an error in the input conformation and can frequently be solved by repairing the input file, instructing Vconf to turn off stereo restraint filters (**-sr**), and/or instructing Vconf not to read chirality information from coordinates (**-cp c**) if they are in 3D. If the input is found to be correct, try increasing the number of searches using the **-ns** option. Errors 10-12 occur when Vconf fails to find stable conformations for a ring fragment. If the input is correct and increasing the number of searches does not help, try repeating the run without searching the rings separately (**-maxra 1**). Errors 13 and 14 occur either when there are no rings or after the rings have been “stitched” together to form the entire molecule.

10. Could not find any ring conformations with energy lower than 1000 kcal/mol with specified cis/trans.

Suggestions:

1. Check the specified isomer. If it is correct, try running with more search steps (**-minrs**).
2. Other stereoisomers may have low energy conformations. Try running with (**-sr n** or **-sr c**). See documentation for details.

12. Could not find any ring conformations with energy lower than 1000 kcal/mol with specified chirality.

Suggestions:

1. Check the specified isomer. If it is correct, try running with more search steps (**-minrs**).
2. Other stereoisomers may have low energy conforma-

tions. Try running with (**-sr n** or **-sr i**). See documentation for details.

12. Could not find any ring conformations with energy lower than 1000 kcal/mol with specified stereochemistry.

Suggestions:

1. Check the specified isomer. If it is correct, try running with more search steps (**-minrs**).
2. Other stereoisomers may have low energy conformations. Try running with **-sr n**. See documentation for details.

13. Could not find any conformations with energy lower than 1000 kcal/mol. Try a larger number of searches [-ns].

14. Could not find any conformations with energy lower than 1000 kcal/mol with specified stereochemistry and cis/trans.

Suggestions:

1. If the specified parities and isomers are correct try running with a larger number of searches [**-ns**].
2. Other stereoisomers or isomers may have low energy conformations. Try running with [**-sr n**, **-sr c** or **-cr i**].

15. Too many atoms

The maximum number of atoms is 999.

16. Molecule setup time limit exceeded

Try increasing the limit with **-ts**.

17. Molecule setup time limit exceeded during resonance calculation.

Try increasing the limit with **-ts** or limit the number of resonance forms generated with **-rl**

18. Molecule setup time limit exceeded during ring fragment preparation.

Try increasing the limit with **-ts** or try **-maxra 15**

Some molecules with very complex rings need more time to find ring conformations. In addition to raising the time-limit, one can also limit ring analysis to rings smaller than the one with the problem. The value after "try **-maxra**" (15 in this example) indicates the number of atoms in the problem ring.

19. Molecule setup time limit exceeded during molecule charging.

Try increasing the limit with **-ts**

The Molecule setup time limit was exceeded during calculation of partial charges.

20. Molecule setup time limit exceeded during molecule typing.

Try increasing the limit with **-ts**

21. There is not enough information given to determine the parity of atom 9; assigning a parity of 3 (undetermined).

For each chiral atom in the molecule, a message is written

in the log file indicating how the parity value was determined: atom parity, stereo bonds, or atomic coordinates. If there was not enough information to determine the parity of an atom (atom 9 in this example), the parity is considered 3; i.e., undetermined.

22.WARNING: Time limit exceeded when filtering conformations. There might be redundant conformations in the output.

Try increasing the limit with **-tf**, or use Vfilter to post-process.

23.The conformation filter failed on this molecule. There might be redundant conformations in the output

COMPUTATIONAL METHODOLOGY

Molecule Setup

Vconf reads and processes all molecule entries in the specified range (*firstMol* to *lastMol*) in the input SDfile. If a molecule entry contains more than one covalently bonded molecule, the one with the most atoms is processed and the others are ignored. For example, if an entry includes a cationic molecule and a chloride ion, the molecule is processed and the chloride is ignored. Vconf command-line options provide the user with considerable flexibility in the interpretation and application of stereochemical information (atom chirality, cis-trans isomers) read from the input file.

Vconf then checks and completes the valences of the atoms. In default mode, this is done by adding missing hydrogens to the molecule, using standard valence rules and formal charge information in the atom block. For example, a hydrogen atom will be added to a nitrogen atom with three explicit substituents and a formal charge of +1. Formal charges are based entirely upon the input file; Vconf does not currently attempt to assign pH-appropriate formal charges. For an input SDfile that already includes hydrogens, it may be more appropriate to use the hydrogens to determine the formal charges. This can be done with the **-fc** option.

Alternate resonance forms [2], but not tautomers, are then identified as required for charging, typing and symmetry detection. The Vcharge method is used to assign VC/2004 charges [2], atoms are typed to the Dreiding force field [4] with minor VeraChem changes to enhance accuracy, and symmetries are detected [3]. The current solvent model is a distance dependent dielectric constant, whose coefficient defaults to 4.0 but can be changed with the **-dc** option.

Rings

Vconf ensures thorough sampling of ring conformations by running a separate conformational search on each ring or ring system in the molecule, and then building conformations of the full molecule from various combinations of the resulting ring conformations. These conformations can then be written out as is (prep mode) or used as starting points for full-molecule conformational searching (search mode). Rings are handled as follows.

Vconf identifies ring systems (solitary rings, as well as more

complex fused and spiro ring systems) within the molecule and splits each such ring fragment out to form a small molecule (a “ring fragment”) ready for conformational searching. The simplest way to form a ring fragment is to use hydrogen atoms fill out the valences of the ring atoms whose bonds were cut to split out the ring, and this option can be invoked with **-sub h**. However, this approach is not recommended because the atoms directly bonded to the ring atoms (the “first level substituents”) can strongly affect the conformational preferences of the ring. For example, a bulky substituent drives the ring’s conformational preferences because it tends to adopt an equatorial position. Therefore, the preferred approach, and Vconf’s default mode, is to include a united-atom representation of the first-level substituents in the ring fragments. For example, a methylene group is represented by a slightly expanded neutral carbon atom. An additional option (**-sub a**) is to use a more complete all-atom representation of the first-level substituents; for example, a methylene group would be represented by a methyl group including 3 explicit hydrogen atoms. This approach works, but does not appear to yield more accurate results than the united-atom method, and it is slower computationally.

Conformation Construction and search

For each ring fragment, the Tork algorithm [1] (see below) is used to identify up to *maxRingConfs* conformations at local energy minima whose energies are within *cutoffRingE* of the most stable conformation found for the fragment. These conformations are stored along with their energies. Full molecule 3D conformations are then constructed by stitching together combinations of the ring conformations. If the set of conformations would lead to greater than 40 full molecule conformations, then the 40 conformations with the lowest sum of ring conformation energies are constructed. Each full molecule conformation is relaxed with a brief torsional Monte Carlo simulation, and then energy-minimized. In prep mode, the resulting conformations are filtered (see below) and written out.

If the construction phase in the previous paragraph yields $N_{\text{initial}} \leq 5$ distinct conformations, all are used as starting conformations for the search phase. If the construction phase yields more than 5 conformations, then the five conformations with lowest total energy are used. The structure that was generated that has the lowest sum of ring energies is not always included in these five conformations. This could happen for two main reasons.

- The set of ring conformations with the lowest energy may force the full molecule to adopt a higher energy conformation.
- The molecule as a whole may, by chance, be assembled in a suboptimal conformation. If one suspects this has happened and that it is adversely affecting the search results, one may use the **-klr** option to ensure that the molecule built from the ring conformations with the lowest sum of energies will be included in the full search.

The search phase starts by subjecting each of these N_{initial} starting conformations to $\text{searchWidth} * \text{numSteps} / N_{\text{initial}}$ steps of con-

formational search with the Tork algorithm (see below), where *searchWidth* is between 0 and 1. The energies of all of the resulting conformations are then compared, and the conformation of lowest energy is subjected to another $(1 - \text{searchWidth}) * \text{numSteps} / N_{\text{initial}}$ Tork steps. Increasing the value of *searchWidth* from 0 toward 1 is considered to increase the width of the search because it causes more Tork steps to be devoted to processing a wide range of initial conformations, rather than to pursuing the lowest energy conformation found when the initial conformations are processed.

Tork Conformational search

Starting from an energy-minimized initial conformation, Tork computes and diagonalizes the second derivative matrix of the energy with respect to all N_{rot} single bonds that are not in a ring fragment. (Using torsional rather than Cartesian coordinates markedly speeds the search because torsional coordinates generate more extensive bond rotations.) The resulting eigenvectors are used to set up N_{rot} “drivers” that smoothly distort the molecule over energy barriers. A single Tork step consists of distortion along a driver, or a randomly selected combination of drivers, followed by energy-minimization to find a local energy minimum. Starting from a given energy minimum, Tork steps are executed until:

- a new local energy minimum (i.e., a more stable conformation) is found, in which case new drivers are constructed and searching continues from the new energy minimum; or,
- all drivers and pairwise combinations of drivers have been used without finding a new conformation of lower energy, in which case the most stable new conformation found during the search is taken as a new starting point; or,
- the requested number of steps (see previous section) has been completed, in which case the Tork search stops.

In the course of all Tork searches, any conformation is immediately discarded if it does not satisfy stereochemical requirements.

Conformational Filtering

Vconf compares all conformations with each other. When two conformations match to within user specified tolerances, the one of higher energy is not written to the output. The filter uses a symmetry detection algorithm [3] originally developed for docking and free energy calculations to make sure that all global and local symmetry operations (such as overall rotation of a molecule, or a benzene flip) are accounted for when comparing two conformations. In addition, alternate resonance forms of molecules are automatically generated in order to avoid missing symmetries that are not apparent from the resonance form of the molecule in the input SDfile. In order to generate greater conformational diversity for molecules with no chiral centers, the program generates the mirror image of every conformation that has survived the conformational filter. If a mirror image is conformationally distinct from the other conformations, according to the filtering criteria, it is included in the output. Therefore it is possible for the file of filtered con-

formations to contain more conformations than the file of unfiltered conformations.

Software Modules

This section lists the sequence of Vconf operations in terms of the software modules that are utilized, in order to help with the interpretation of error messages and warnings.

1. Setup Module
 - a. Check that molecule has atoms.
 - b. Check for radicals.
 - c. Check for unspecified bond orders.
 - d. Break artificial covalent bonds linking counterions to main compound in some SDfiles. (Only breaks bonds to K and Na.)
 - e. Delete all chemical fragments but the largest; e.g., deletes counterions without covalent bonds.
 - f. Standardize nitro notation: O=N=O becomes O=N-O(-1).
 - g. Either add hydrogens to fill valences based on formal charges in SDfile (default), or assume all hydrogens are present and assign corresponding formal charges.
 - h. Check valences of prepared molecule.
 - i. Interpret stereochemical specifications in SDfile.
2. Force Field Parameterization
Atom Charging Module and Force Field Typing Module use Ring Finder, Aromaticity Finder and Resonance Generator modules.
3. Ring Preparation Module
 - a. Identify ring fragments in molecule and build small molecules based on them, using the selected substituent style.
 - b. Generate ring conformations from each fragment using Tork and conformational filter.
4. Conformational Search Module.
 - a. If molecule contains ring fragment(s), build starting conformations of the full molecule from ring conformations and run Tork conformational search. Otherwise, simply run Tork. During Tork calculations, delete conformations that violate specified stereochemistry.
 - b. Filter conformations, using symmetry information, to remove repeats.

NOTES

1. Vconf can process an SDfile in which all molecules have formal charges specified and hydrogens need to be added, or an SDfile in which all molecules have hydrogens and formal charges can be deduced. However, if an SDfile mixes these specifications, then some molecules will not be processed. The default is to assume formal charges are correct and add hydrogens as needed. Use the **-fc** option in Command Line Options if all hydrogens are already present and you want Vconf to assign formal charges.
2. The names of the output SDfiles with the conformations generated for each molecule in the input SDfile are based on molecule names (*SDfilename_confs.sdf*). Therefore, it is

helpful if the molecule names in the name fields of the SDfile or Molfile are unique.

TROUBLESHOOTING

1. If Vconf does not provide as many conformations as expected, try any or all of the following:
 - a. Increase the number of searches, numSteps.
 - b. Save the unfiltered conformations (**-u**) and use Vfilter (available separately) to filter with less stringent criteria.
 - c. Generate more ring conformations (use **-maxrc**).
 - d. Rerun with different random number seeds, merge the outputs of the two runs, and refilter using Vfilter. Note that a new run will automatically use new random number seeds unless the seeds are specified in the command line.
2. If Vconf fails to find a conformation with the desired stereochemistry and increasing the number of searches does not help:
 - a. Do the run with a different set of random number seeds (see 1d above).
 - b. If the problem occurs during generation of ring conformations (see error messages in log file), try generating the ring's conformations during the full-molecule search by reducing the maximal ring size with the **-maxra** *maxRingAtoms* option so that the ring is no longer treated as a fragment.
3. If more than 40 ring conformations are desired, reduce maxRingAtoms with the **"-maxra"** option. This will cause Vconf to search the ring during the full-molecule part of the search as in 2b above.
4. E-mail other problems to support@verachem.com. Please include the input and output files, along with a description of the problem.

REFERENCES

1. Tork: A conformational analysis method for molecules and complexes. Chang, C. & Gilson, M.K., *J. Comput. Chem.*, 24:1987-1998, **2003**.
2. Fast assignment of accurate partial atomic charges. An electronegativity equalization method that accounts for alternate resonance forms. Gilson, M.K., Gilson, H.S.R. & Potter, M.J., *J. Chem. Inf. Comput. Sci.*, 43:1982-1997, **2003**.
3. Identification of symmetries in molecules and complexes. Chen, W., Huang, J., & Gilson, M.K. *J. Chem. Inf. Comput. Sci.* 44:1301-1313, **2004**.
4. A generic force field for molecular simulations. Mayo, S. L., Olafson, B. D., Goddard III, W. A. Dreiding, *J. Phys. Chem.* 94:8897-8909, **1990**.
5. Development and testing of a general AMBER force field. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A., *J. Comput. Chem.*, 25:1157-1174, **2004**.
6. Automatic atom type and bond type perception in molecular mechanical calculations. Wang, J., Wang, W., Kollman P. A., & Case, D. A., *J. Mol. Graph. and Model.*, 25:247-260, **2006**.

Notice: The Vconf software, including executables, source code, this documentation, and other associated files, are protected by copyright. The software is also protected by patents pending. Usage of all these materials is governed by a license agreement with VeraChem LLC, P.O. Box 2206, Germantown, MD 20875.