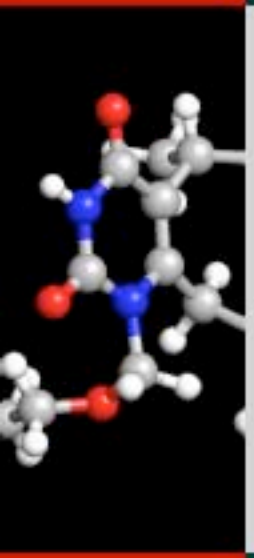


Conformational Searching using MacroModel and ConfGen

John Shelley
Schrödinger Fellow

Overview

- **Types of conformational searching applications**
- **MacroModel's conformation generation procedure**
- **General features supporting conformational searches**
- **Conformation generation methods**
  - Methods which accumulate changes
  - Methods which exhaustively sample local minima
- **General guidelines**
- **Closing points**

Types of conformational searching applications

Types of systems

- Ligands
- Proteins and protein-ligand complexes
- Other

Characteristics

- Size of molecules
- Scale of conformational changes expected
- Thoroughness
- Collectivity of conformational changes
- Time/disk space usage

MacroModel's conformation generation procedure

.....For Each molecule.....

1. **Read the input structure**
2. **Characterize the molecule**
   - Stereochemistry: chiralities and double bond geometries (E/Z)
   - Molecule/search method specific features*
     - Symmetry
     - Comparison atoms
     - Rotatable bonds
     - Ring systems (and ring opening bonds)
     - Restraints
     - Number of steps
     - Etc.
3. **Optionally energy minimize the input structure**
4. **Generate conformers one at a time**
   - Generate a new conformation for the molecule*
   - Optionally post process the conformation
     - Estimate energy or minimize
     - Filter by energy
     - Eliminate redundant conformers
4. **Save results**

* These vary with the search method employed

General features supporting conformational searches

1. **Force fields (all atom)**
   - OPLS_2005, OPLS_2001
   - MMFF, MMFFs
2. **Solvent models**
   - Vacuum (or constant dielectric)
   - Distance dependent dielectric (1/4r)
   - GB/SA
3. **para_bmin**
   - For distributing the separate searches of different molecules across multiple processors
   - Fault tolerant
4. **dbmin**
   - For distributing the search of a single structure across multiple processors
   - Useful for protein-ligand complexes.

Conformation generation methods that accumulate changes

Methods which accumulate changes

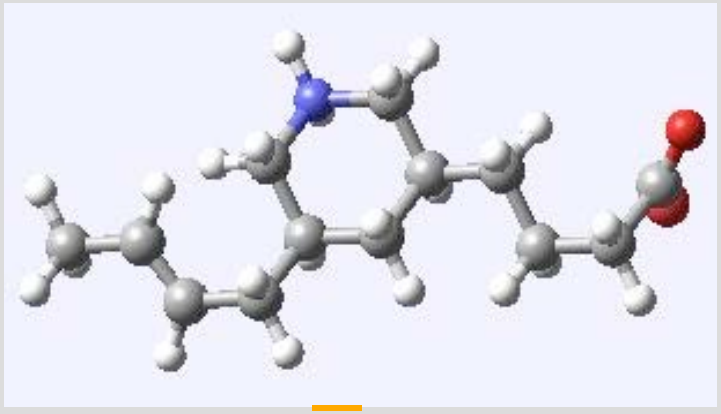
A collection of conformations is built up during the search.

Conformations generated earlier in the search are used as starting points for subsequent search steps.

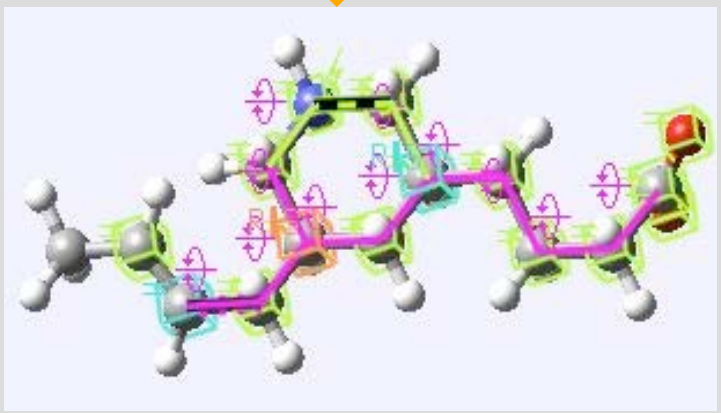
Methods:

1. MCMM, Monte Carlo Multiple Minimum
   - Random torsional search plus minimization
2. Low MODe (LMOD) and Large-scale Low Mode (LLMOD)
   - Use low frequency modes to construct conformational changes
3. Mixed Mode
   - A combination of MCMM with one of LMOD or LLMOD

MCMM (Monte Carlo Multiple Minimum)



Automatic Setup



Each MCMM step consists of:

1. **Select an existing conformer**
2. **Introduce changes by randomly:**
   - adjusting the dihedral angles for rotatable bonds
   - Translating and rotating molecules (if there are two or more molecules)
3. **Sampling rings:**
   - Pretend to break a bond in the ring (the ring opening bond)
   - Treat the remaining single bonds as rotatable.
   - If the atoms in the broken bond end up lying close to each other reform the bond. Otherwise, try another set of bond rotations
4. **Post process the structure:**
   - Minimization
   - Stereochemical checks
   - Energy window
   - Redundancy check
   - If OK, add structure to collection of conformers

MCMM: Strengths and Weaknesses

Strengths

1. General method which can be applied to both small and large structures
2. Very effective if rotations about a small number of bonds will yield a large unhindered conformational change

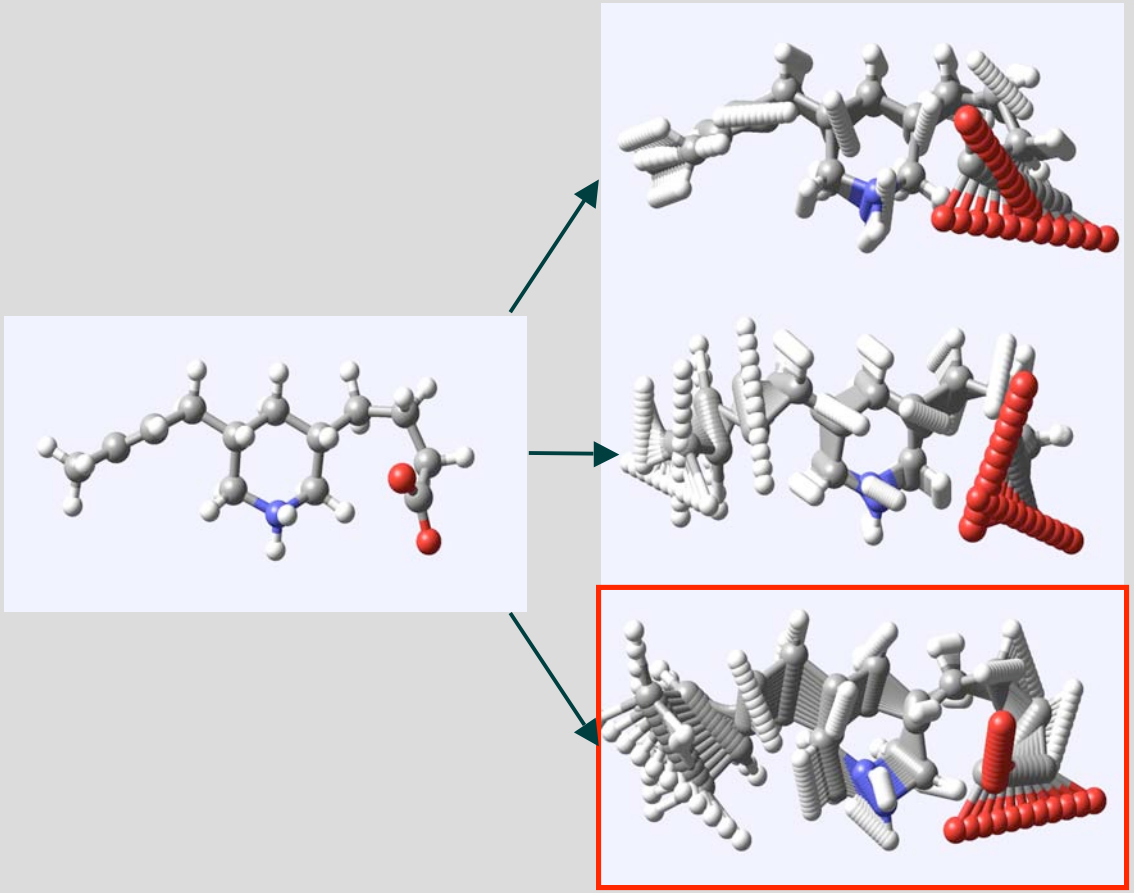
Weaknesses

1. Need to limit the number of simultaneous rotations
2. Large conformational changes may need a series of localized changes. Searches can be long and CPU-intensive.
3. Not so good for environments in which explicit atoms fill up most of the space because changes need to be collective.

Low Mode and Large scale Low Mode

For each LMOD or LLMOD step:

1. **Select an existing conformer**
2. **Generate normal modes**
3. **Select a normal mode**
   - Amplify the normal mode so as to generate a very different structure
4. **Post process the structure**
   - Minimization
   - Stereochemical checks
   - Energy window
   - Redundancy check
   - If OK, add structure to collection of conformers



LMOD and LLMOD: Strengths and Weaknesses

Strengths

1. General method which can be applied to both small and large structures
2. Useful for environments in which explicit atoms fill up most of the space causing changes to be collective

Weaknesses

1. Large conformational changes may need a series of localized changes (more so than MCMM). Searches can be long and CPU-intensive.
2. Individual search steps use more computer time than MCMM.
3. Sometimes this is not effective at sampling translations and rotations of molecules in complexes.

Mixed mode searches

Idea

Use the strengths of both methods by performing a search in which some steps are LMOD or LLMOD steps and others are MCMM steps.

Implementation

Just like either of these search methods, but at each step randomly choose which type of step to use

Can set the probability MCMM vs LMOD or LLMOD (usually 50/50 works fine)

Mixed mode searches are recommended over searches employing one search technique.

Exploration: Can we also usefully employ fairly new methods like FLAP (ring corner reflection) and MCRC (ring sampling using pre-generated conformations) in mixed mode searches?

GUI for methods which accumulate changes

Accessing

Maestro

Applications (menu)

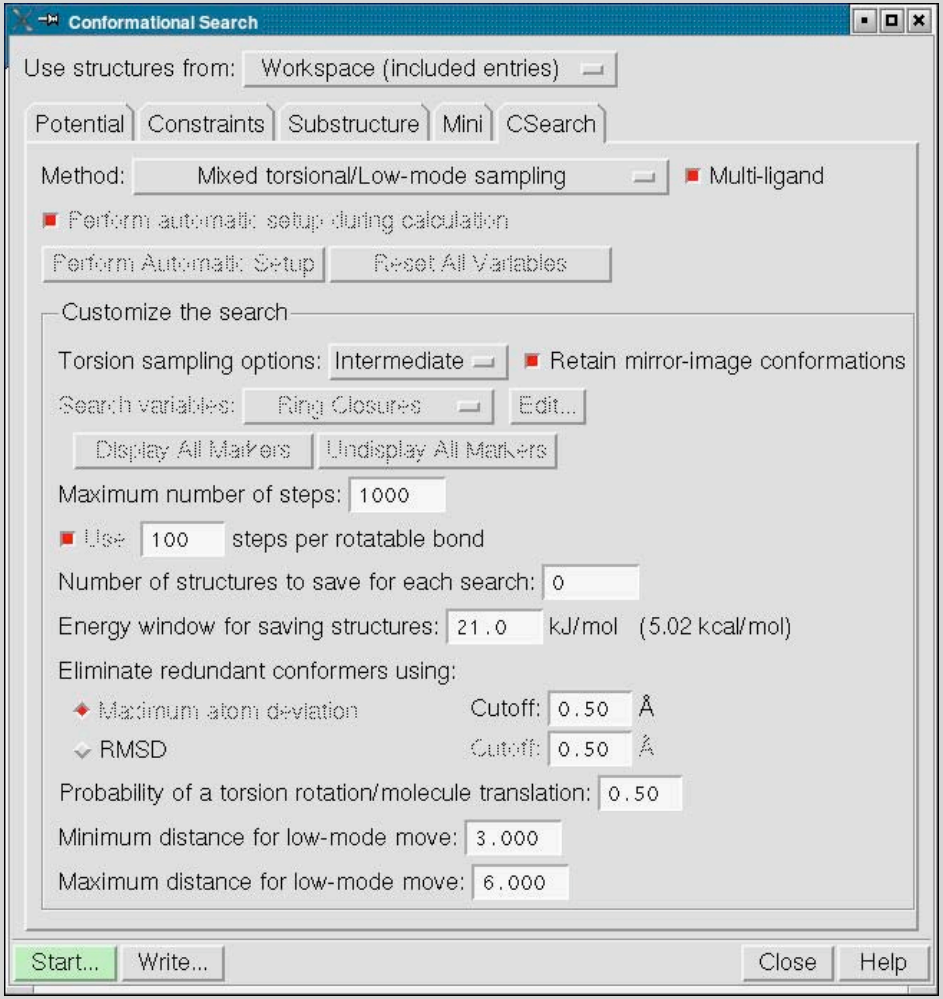
MacroModel

Conformational Search…

CSearch (tab)

Recommendation:

For ligands, the default settings are appropriate.

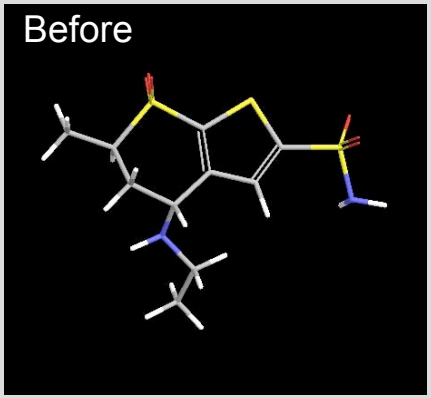


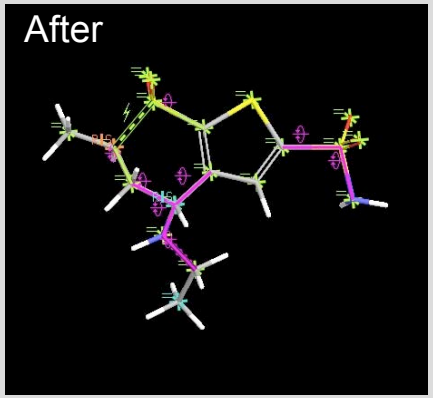
Automatic setup

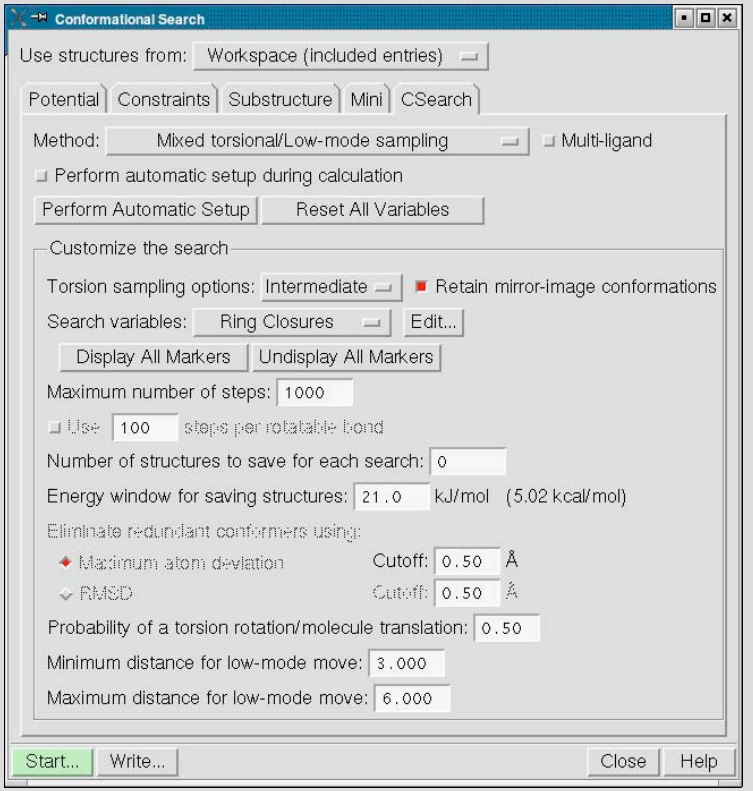
What is automatic setup doing?

To find out:

1. Place an interesting molecule in the workspace
2. Turn off the `Multi-ligand toggle`
3. Turn off `Perform automatic setup during calculation`
4. Hit `Perform Automatic Setup`

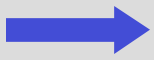


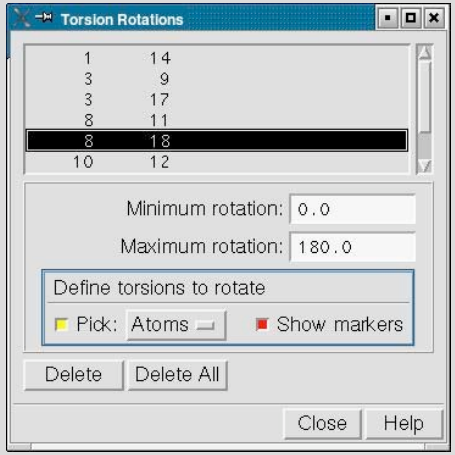




How can I modify the setup?

Most features can be changed using the `Search variables` pull-down menu and then hitting the `Edit...` button.





Hints: mixed mode searches of protein-ligand complexes

1. **Be selective about what bonds to rotate using MCMM (see previous slide)**
   - Typically only a few such rotations can be changed at one simultaneously.
   - Focus those changes on key rotatable bonds that look like they can have a significant chance of resulting in larger changes.

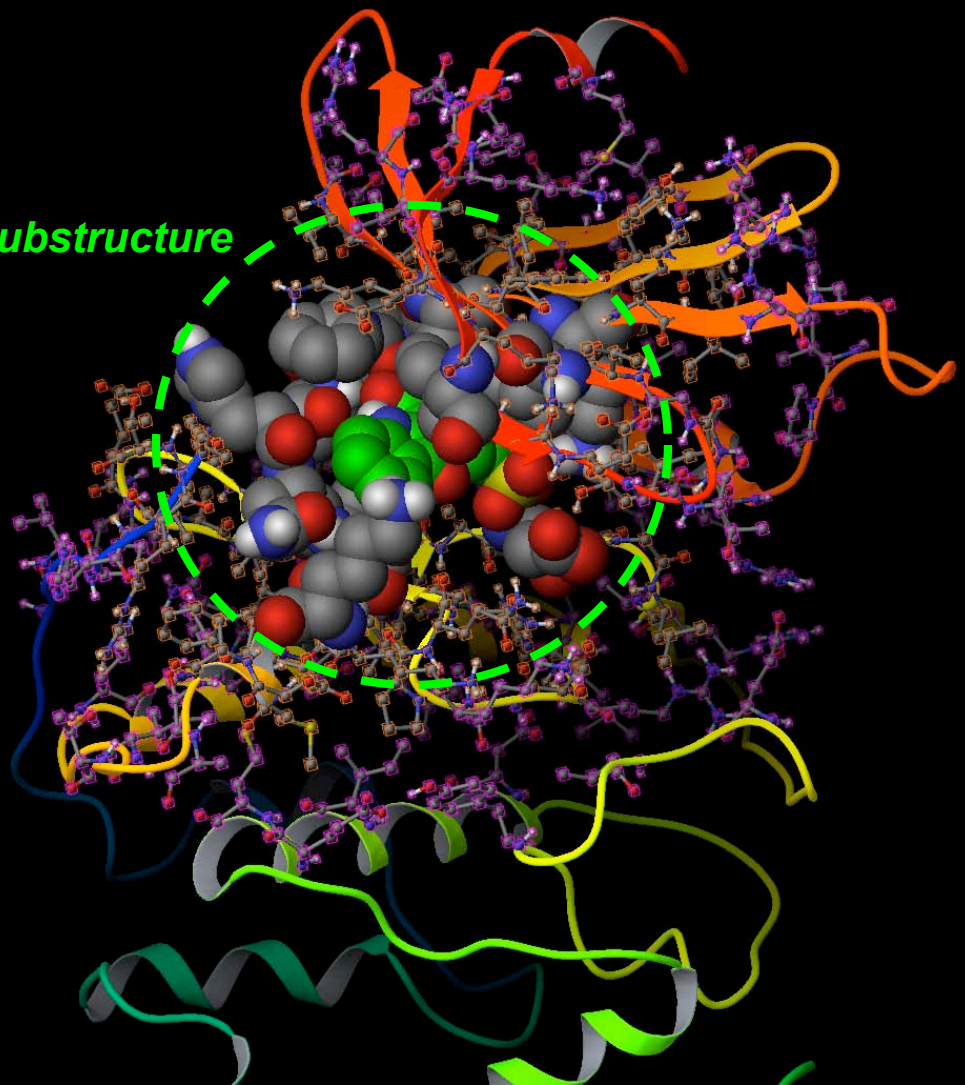
2. **Use substructures (see next slide)**

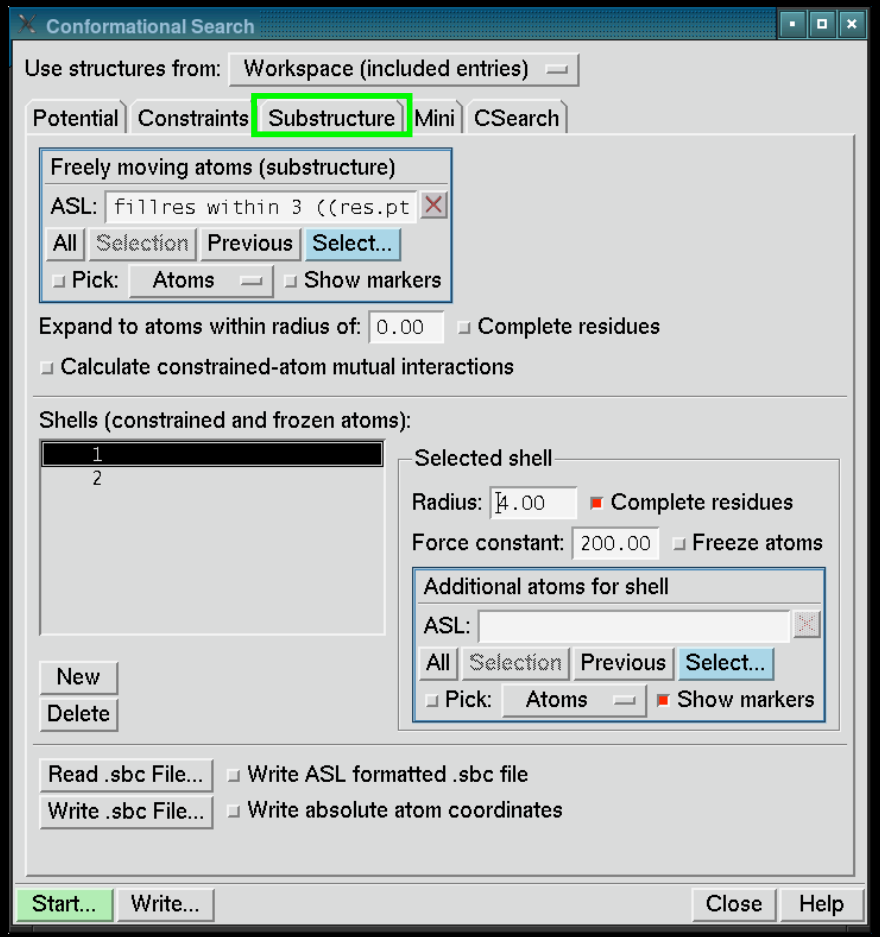
Defining a substructure region for conformational sampling

Substructure – flexible atoms

Shell 1 – constrained atoms

Shell 2 – frozen atoms





Options:

- Write .sbc file for a future use
- ASL formatted .sbc file

Conformation generation methods that exhaustively sample local minima

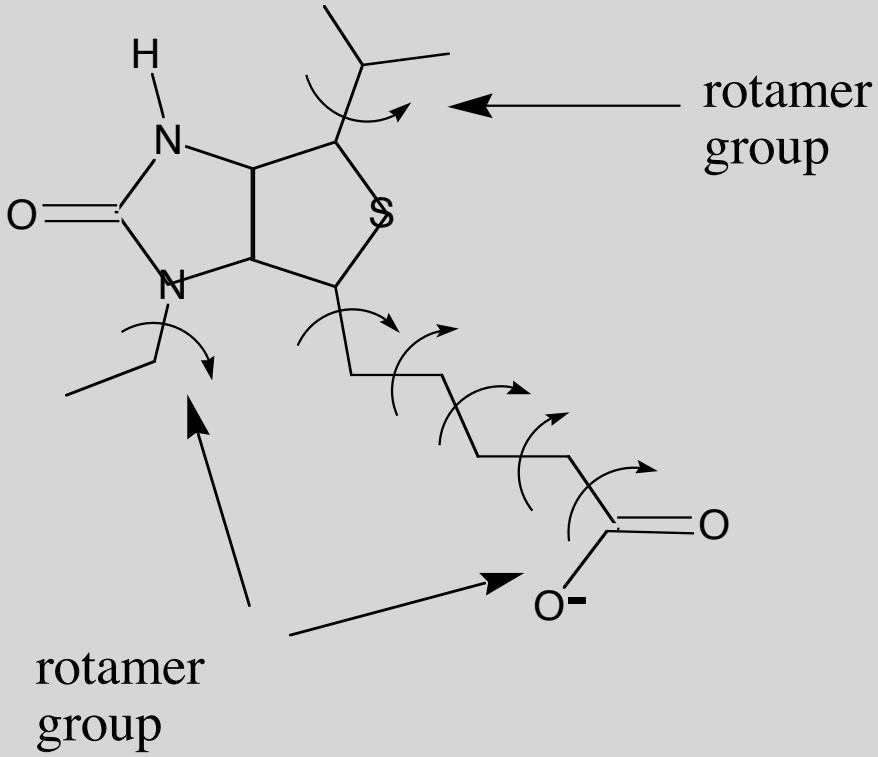
Methods which exhaustively sample local minima - ConfGen

Overall goal: systematically explore all or nearly all conformations

Key features:

1. Identify low energy ring conformations
   - I. Flexible ring systems are identified
     - Mutually connected flexible rings and aromatic rings that share at least two atoms with one of the flexible rings in the ring system
   - II. A corresponding template is sought from a large template library
     - i. Template matching is strict; element and hybridization for all atoms in the template must match the target ring system.
     - ii. Low energy conformers for each template ring system have already been generated and stored in the template library
2. Identify invertible nitrogen atoms
   - Non-ring pyramidal sp3 nitrogen atoms that are bonded to only three other atoms (i.e., unprotonated sp3 nitrogen – not very common)
3. Identify rotatable bonds
   - Single bonds whose rotations result in non-trivial changes in the molecules

ConfGen: Concept of Molecular core



ConfGen: Core geometry sampling

For all combinations of low energy ring system conformations and nitrogen atom inversions:

1. **Identify minima for rotating about each rotatable bond**
   - I. **Generate tabulated potentials for each rotatable bonds**
     - i. Use OPLS_2001
       - dihedral
       - Lennard-Jones non-bonded interactions for a subset of the atoms on either side of the bond that are deemed key to avoiding topologically local vdw clashes
     - ii. Enforce rotational symmetry
     - iii. Option: If the range of the force field based potential is small replace it with a cosine potential (typically 6 fold)
   - II. **Identify minima in the tabulated potentials**
     - i. Limit to at most 10 minima per rotatable bond
     - ii. Eliminate equivalent minima for bonds with rotationally symmetry.
     - iii. Determine the relative energies of the minima
2. **Generate conformations for all combinations of minima for rotatable bonds within the core**
   - I. **Eliminate high energy conformations on the fly based upon relative ring energies and the relative energies of the minima for the rotatable bonds.**
   - II. **Minimize the energy by adjusting the dihedral angles for the rotatable bonds.**
   - III. **Score core geometries and prune by score if excessive core conformations are produced.**

ConfGen: Rotamer geometry sampling

For each core geometry sample rotamers of the peripheral groups:

1. **Use one of two user selectable approaches:**
   - Sample all combinations of peripheral group rotamers
   - Excite each peripheral group in turn while keeping the other groups in their lowest energy minima.
2. **Eliminate high energy conformations (as for core conformation sampling)**

ConfGen: Conformation filtering

Goal: Eliminate undesirable conformations

Filter out conformations with:

1. **vdW overlap**
   I. Conformations in which heavy atoms approach each other more closely than 60% of the sum of their vdw radii are eliminated.
   II. If the previous step eliminates all of the conformations then reduce the closest approach distance by 15% and try again.
   III. If the previous step eliminates all of the conformations then reduce the closest approach distance by another 15% and try again (this is rarely needed).
2. **Close approach of atoms with formal charges of the same sign**
3. **A hydrogen bonds pointing towards nearby atoms that have a net positive formal charge**
4. **high local concentrations of heavy atoms (e.g. stacked rings)**

ConfGen: Selecting a subset of conformers

Limit the number of conformers to at most:

$N_{target} = (N_{mult})$(Effective number of degrees of freedom)

N_{mult} is a parameter

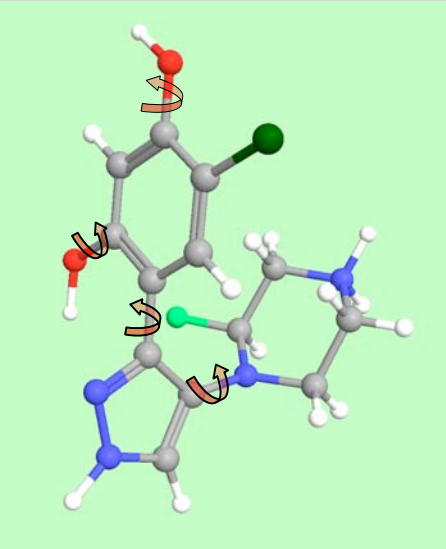
Uniformly selecting a subset of all of the conformers

Conformers ranked by score

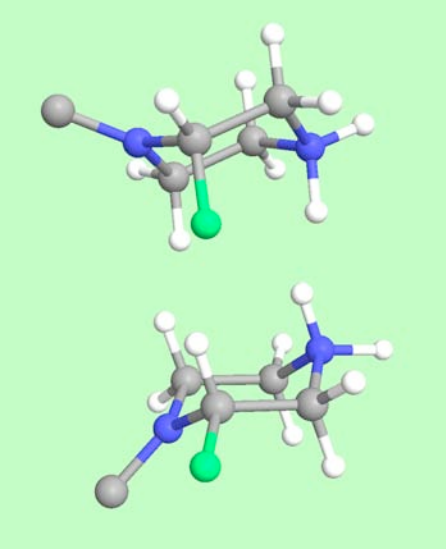
1. Select the first conformer
2. Select the last conformer
3. Select conformers ½ way between those already selected in the ordered list of conformers
4. Repeat 3 until sufficient conformers have been selected

ConfGen: conformation generation

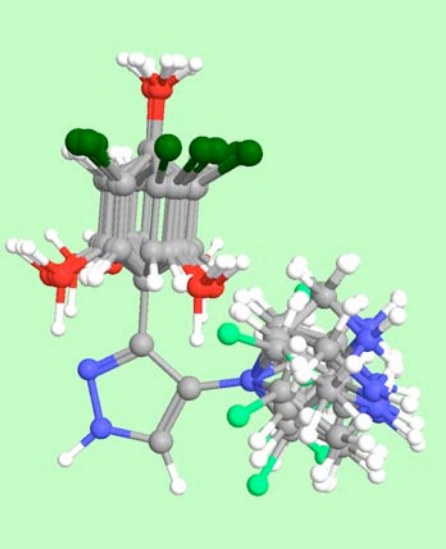
Sample Rotatable Bonds



Sample Ring Conformations



Resulting Conformations



ConfGen and MacroModel

ConfGen functionality was originally developed for Glide and has undergone further development for use with other products.

ConfGen searches can be performed using Phase or MacroModel. ConfGen/MacroModel calculations:

- are supported by ConfGen Panels in Maestro
- permit MacroModel's pre- and post-processing functionalities to work synergistically with ConfGen's. Namely:
  - the use of a range of force fields
  - solvation models
  - minimization
  - filtering (energy, redundancy)
- Net result is a reduction in the number of conformers along with an improvement in the quality of the conformations.

ConfGen/MacroModel search optimization study

Ligands Used: All studies were conducted on a set of 667 ligands, including 36 from Boström (Comp. Aided Mol. Design 2001, 15, 1137-1152) , the 100 public structures from Perola and Charifson (J. Med. Chem. 2004, 47, 2499-2510), and an additional 538 selected from the Protein Data Bank

Defined 4 protocols

Very fast

- no energy evaluations (or minimizations) in MacroModel
- RMS for redundant conformer elimination: 1.25
- $N_{mult} = 5$

Fast

- Eliminate conformers with high all atom relative energies > 12 kJ/mol
- RMS for redundant conformer elimination: 1.0
- $N_{mult} = 5$

Intermediate

- Eliminate conformers with high all atom relative energies > 104.7 kJ/mol
- RMS for redundant conformer elimination: 1.0
- $N_{mult} = 75$

Comprehensive

- Eliminate conformers with high all atom relative energies > 500 kJ/mol
- RMS for redundant conformer elimination: 0.5
- $N_{mult} = 75$

ConfGen/MacroModel search optimization study

Systematic Exploration of how to use ConfGen and MacroModel together

Method	Root Mean Square distance in Å*				# of	Time**
	<0.5	<1.0	< 1.5	<2.0	Confs	
Very Fast	16%	52%	84%	96%	14.6	1.0
Fast	19%	55%	82%	95%	13.5	2.3
Intermediate	21%	64%	85%	97%	38.4	6.5
Comprehensive	34%	70%	90%	98%	150.2	16.3

* Percent of ligands with a generated conformer with an RMS relative to the bioactive conformer less than value listed

**Time in seconds on a AMD Opteron 1.6 GHz processors with 2 GB memory.

ConfGen/MacroModel search optimization study

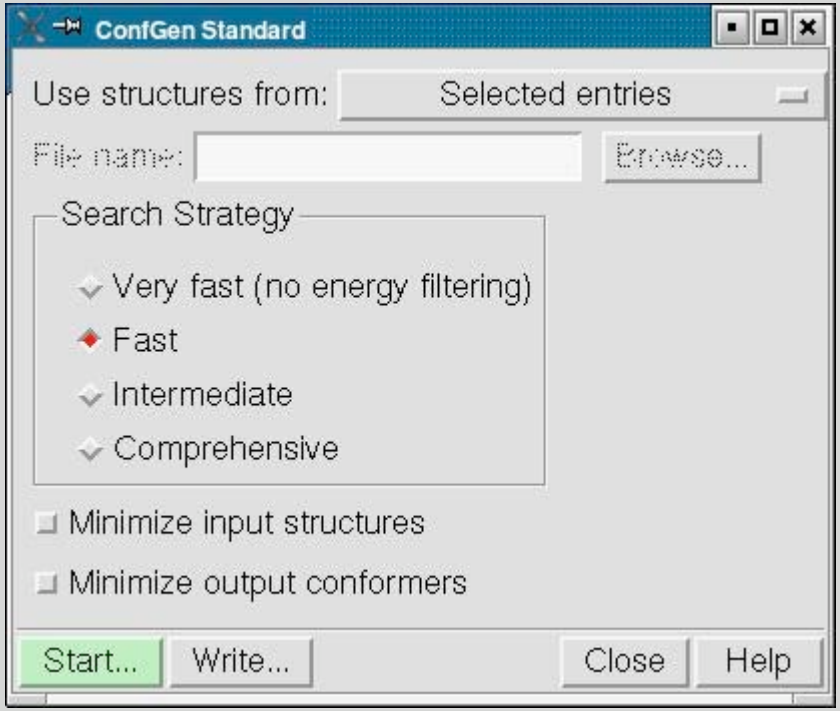
- Comprehensive has the best results at all levels of accuracy.
- At the standard level used in the literature for matching bioactive conformations (<2.0 Å) the coverage is nearly complete for all 4 protocols.
- Searches are fast enough for use in generating databases of conformers
- Very fast, Fast, and Intermediate produce fewer conformers than is typically used/required (approximately 100-200) for this level of coverage at (<2.0 Å).
  - Lower disk space usage for large databases
  - Computations using these conformers should use less CPU time.
  - Fewer irrelevant and perhaps high energy conformers that may generate spurious results in subsequent processing (e.g. pharmacophore generation)
- These are done without minimization. Minimization generally increases CPU time by a factor of 4-12, while increasing the matches by 1-5 percent.
- Versatile with a broad range of speed/quality trade offs.

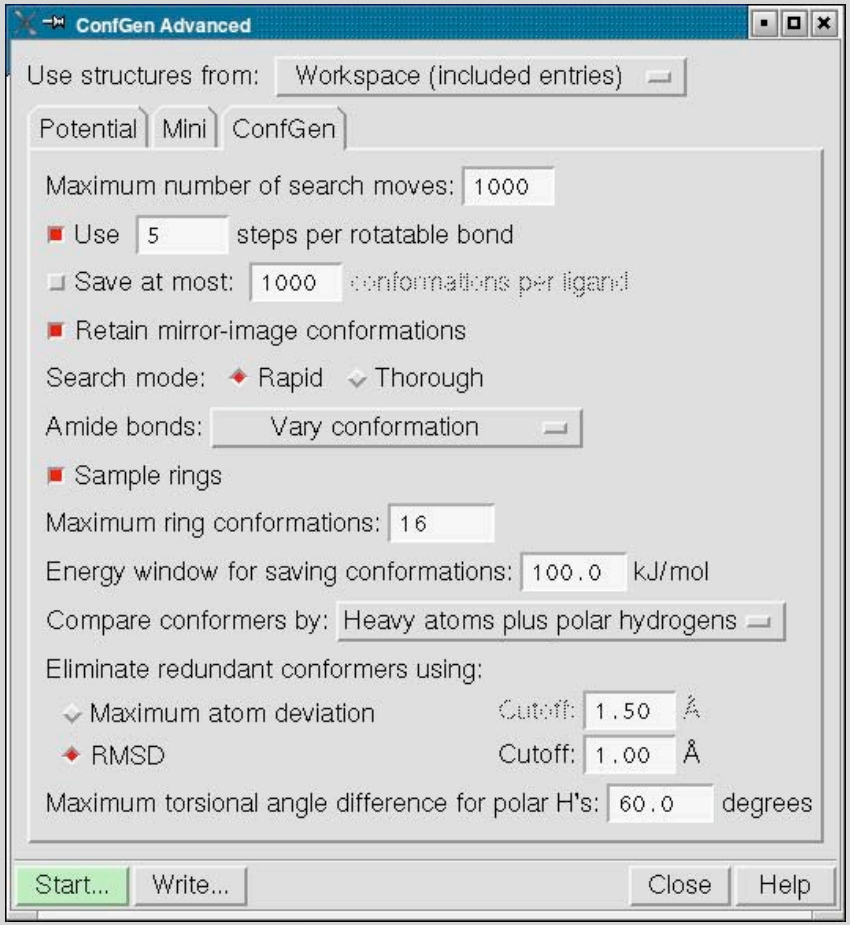
ConfGen GUI

Accessing:

- **Maestro**
  - **Applications (menu)**
    - **ConfGen**
      - **Standard or Advanced**

***Recommended ***





ConfGen: Strengths and Weaknesses

Strengths

1. Fast enough to be used for large databases of ligands
2. Exhaustive sampling of local minima ensures comprehensive coverage of conformer space
3. Small subsets of conformers provide good coverage of conformer space

Weaknesses

1. Limited to ligand-sized molecules
2. Even exhaustive sampling of local minima may miss some minima that arise from non-local interactions.

General Guidelines

For Ligands

- Fairly accurate yet quite fast (seconds/ligand) – ConfGen without minimization
  - e.g. conformer database generation
- Quite accurate yet fairly fast (1 minute/ligand) – ConfGen with minimization
- Very accurate yet slow (1 hr/ ligand)– MCMM/LMOD mixed searches
  - e.g. generating conformers for using in creating pharmacophores

Protein/Ligand complexes

- Use substructures
- Use mixed mode MCMM/LLMOD or MCMM/LMOD (< 250 moving atoms) searches
- Consider selecting MCMM torsional moves carefully

Other systems

- **Approach usually needs to be tailored to the nature of the problem**
- **For large system mixed mode MCMM/LLMOD is usually the appropriate choice**
  - Use larger ranges for LLMOD moves
  - Careful selection of MCMM torsional moves
  - Use larger energy window (reminimize output later)
  - Increase RMS values somewhat

Closing points

- Together ConfGen and MacroModel can be used to search a wide range of molecules from ligands to protein/ligand complexes and other types of molecules.
- Maestro makes setting up and running these calculations easy.
- Wide range of speed/quality trade offs possible.
- Small subsets of conformers generated by ConfGen can cover conformational space quite well.
- Searching multiple structures in a single calculation automatically is supported.
- Distributing calculations is supported.

Acknowledgments

MacroModel team

dozens of people over the years

ConfGen Development and Application studies

Pranav Dalal
Rich Friesner
Robert Murphy
Woody Sherman
Shawn Watts