DEALING WITH CONFORMATIONALLY FLEXIBLE MOLECULES

Very few molecules are rigid and adequately described in terms of a single conformer. More commonly, two or more distinct conformers resulting from rotation around single bonds or changes in flexible rings are likely to exist. For example and as previously discussed in the topic **Potential Energy Surfaces**, three conformers result from rotation around the central carbon-carbon bond in *n*-butane, one *anti* conformer (CCCC dihedral angle = 180°) and two (equivalent) *gauche* conformers (CCCC dihedral angle = ± ~60°). Two chair conformers result from ring inversion in methylcyclohexane, one with the methyl group *equatorial* and the other with it *axial*. Because energy barriers to single-bond rotation and to ring inversion are generally very small (on the order of <10-20 kJ/mol), the different conformers will be in equilibrium and the influence of any particular conformer on the properties of the molecule will depend on its energy and on the temperature. As a consequence, experimental measurements will either reveal the presence of the individual conformers or result from a weighted average of conformers. For example, the IR spectrum of a flexible molecule will show features due to individual conformers, whereas the NMR spectrum of the same molecule will show an average. The different behavior may be traced to the different timescales of the two experiments; IR is fast whereas NMR is slow.

Identifying the lowest-energy conformer of a molecule with multiple degrees of conformational freedom presents a serious challenge for calculation. Even more so is providing accurate Boltzmann weights allowing proper description of average energies, properties and perhaps most important, NMR spectra. Until recently, molecular mechanics provided the only practical alternative for molecules with hundreds or even thousands of accessible conformers. Expectations were high for molecular mechanics force fields such as MMFF which were after all developed and parameterized to account for conformer energetics, albeit for very small molecules with a single degree of conformational freedom for which experimental data was available.[*]

---

[*] While much is known about the preferred conformation of molecules with multiple degrees
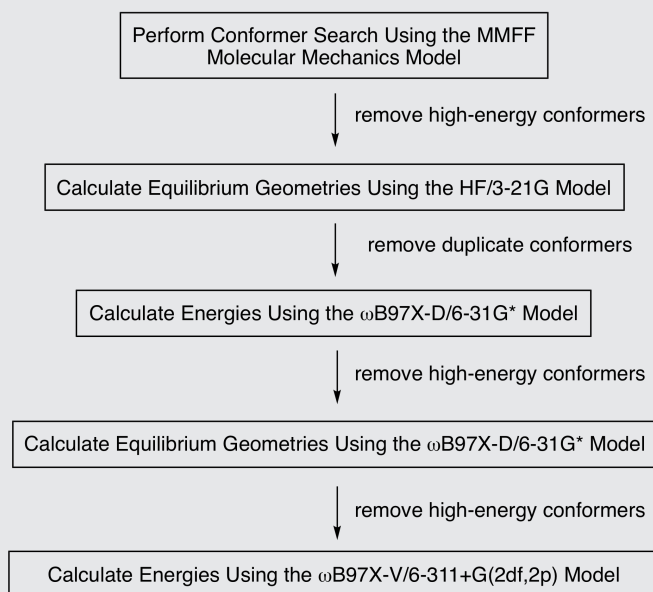
However, recent comparisons with conformer energies obtained from high-level quantum chemical calculations for (very few) molecules with multiple degrees of freedom show that MMFF actually provides a poor account. Furthermore, it has been shown that neither semi-empirical nor Hartree-Fock models provides a suitable account and that either density functional or MP2 models with moderate to large basis sets are required.

Direct application of correlated models is not presently practical or at least routine for molecules where hundreds or more conformers need to be examined. What is practical are multi-step procedures starting from an extensive molecular mechanics search and ending with an appropriate model for energy calculation. The idea is that intermediate steps with successively better models will gradually reduce the number of conformers that need to be considered (passed on to the next step). Several present recipes (available from menus) are supported in *Spartan*. The simplest (least costly and likely least reliable) use either B3LYP/6-31G* or ωB97X-D in the final step and MMFF equilibrium geometries. Improvement (and increased computation cost) follows by using B3LYP/6-31G* or ωB97X-D geometries. The most reliable (and most costly) models use B97M-V/6-311+G(2df,2p)[6-311G*] or ωB97X-V/6-311+G(2df,2p)[6-311G*] for final energy (Boltzmann weight) calculations together with B3LYP/6-31G* or ωB97X-D/6-31G* geometries, respectively. Whatever the recipe employed, the resulting Boltzmann weights are available to determine average quantities.

---

of conformational freedom in the solid (crystalline) state, almost exclusively from X-ray diffraction, there is very little known about preferred conformation of isolated molecules. Furthermore, there is very little overlap between gas and solid-phase structures because small molecules tend not to crystallize and larger molecules that do crystallize are not easily investigated by gas-phase techniques such as microwave spectroscopy. This means that it is difficult to separate inherent (gas-phase) conformational preferences from those associated with the requirements of crystal packing.

One of the menu-selected multi-step protocols is outlined below.

```
┌─────────────────────────────────────────┐
│  Perform Conformer Search Using the MMFF │
│        Molecular Mechanics Model         │
└─────────────────────────────────────────┘
                    │ remove high-energy conformers
                    ▼
┌─────────────────────────────────────────┐
│ Calculate Equilibrium Geometries Using the HF/3-21G Model │
└─────────────────────────────────────────┘
                    │ remove duplicate conformers
                    ▼
┌─────────────────────────────────────────┐
│ Calculate Energies Using the ωB97X-D/6-31G* Model │
└─────────────────────────────────────────┘
                    │ remove high-energy conformers
                    ▼
┌─────────────────────────────────────────┐
│ Calculate Equilibrium Geometries Using the ωB97X-D/6-31G* Model │
└─────────────────────────────────────────┘
                    │ remove high-energy conformers
                    ▼
┌─────────────────────────────────────────┐
│ Calculate Energies Using the ωB97X-V/6-311+G(2df,2p) Model │
└─────────────────────────────────────────┘
```

Step 1: Conformer Search Using the MMFF Molecular Mechanics Model

The first and least costly step in the protocol involves a systematic conformational search using the MMFF molecular mechanics model, keeping all distinct conformers with energies less than 40 kJ/mol than that of the lowest-energy conformer found. The energy threshold is needed to control the number of conformers passed on to the next step, recognizing that a molecule with 8 degrees of conformational freedom may give rise to several thousand distinct conformers and one with 10 degrees of freedom to several tens of thousands of conformers. In addition, a limit on the total number of conformers needs to be imposed. If exceeded, a subset of the most "geometrically diverse" set is kept. This number is typically in the range of 200-1000 conformers.

Step 2: Calculate Conformer Geometries Using the HF/3-21G Model

Obtain the equilibrium geometry for each of the conformers using the Hartree-Fock method with the 3-21G basis set (HF/3-21G or simply 3-21G). Eliminate any conformers that are outside a 40 kJ/mol threshold as well as duplicate conformers. This step is necessary because MMFF commonly results in structures that are not energy minima, and further optimization using the HF/3-21G model allows these to be removed. Note, however, that this model is not able to reliably identify the "best" conformer let alone provide accurate Boltzmann weights, which is why

the 40 kJ/mol energy threshold needs to be maintained.

Step 3: Calculate Conformer Energies Using the ωB97X-D/6-31G* Model

Use the ωB97X-D/6-31G* model to calculate energies for each of the conformers, keeping only those that are within 15 kJ/mol of the lowest energy conformer. The much lower energy threshold reflects experience that, even with a small basis set, the ωB97X-D model provides conformer energy differences that are much more reliable than those from other MMFF or HF/3-21G models.

Step 4: Calculate Conformer Geometries Using the ωB97X-D/6-31G* Model

Recalculate equilibrium geometries using the ωB97X-D/6-31G* model, keeping only conformers that are within 10 kJ/mol of the lowest energy conformer. The slight lowering of the energy threshold anticipates that better equlibrium geometries will lead to more reliable conformer energy differences.

Step 5: Calculate Conformer Energies Using the ωB97X-V/ 6-311+G(2df,2p)[6-311G*] Model

The final step in the protocol is to replace ωB97X-D/6-31G* conformer energies with those obtained from the ωB97X-V/6-311+G(2df,2p) model. In practice, the so-called dual basis set approximation is employed, reducing the cost of this step by as much as an order of magnitude.

In terms of computation, the last two steps in the protocol outlined above are the most costly: calculation of equilibrium geometries with the ωB97X-D/6-31G* model and calculation of energies with the ωB97X-V/6-311+G(2df,2p)[6-311G*] model.

A less costly (factor of two) protocol has also been examined and found to produce comparable albeit less reliable results. This replaces ωB97X-D/6-31G* by B3LYP/6-31G* and ωB97X-V/6-6311+G(2df,2p)[6-311G**] by B97M-V/6-311+G(2df,2p)[6-311G*].

Where only the identity of the lowest-energy conformer is required, **_Spartan_** employs the so-called **_Monte Carlo_** procedure. It moves randomly in a single (randomly chosen) dimension in conformational space deciding to abandon the move (returning to the starting point for another random move) or keep the move (using it for the starting point for the next random move) based on the energy relative to that of the

lowest-energy conformer yet found. In practice, a Monte-Carlo search will nearly always find the lowest-energy conformer by considering only 5-10% of the possible number of conformers.

A Monte-Carlo search is probably not an appropriate starting point where the objective is to establish a Boltzmann distribution. The reason is that it rapidly finds the "best" region, at the expense of other regions, which, given the known deficiencies of molecular mechanics, may exclude the actual important contributors (or even the actual most important contributor). The only alternative is a ***systematic search*** which "looks everywhere". However, this rapidly becomes impractical as the number of conformers increases very rapidly with the number of rotatable single bonds and flexible rings. The solution employed in ***Spartan*** is a so-called ***sparse systematic search*** where systematic steps are accepted (and energies evaluated) or rejected based solely on a random number criterion. For example, if at the outset only 10% of the possible conformers can be examined, then only those generating a random number between say 0 and 0.1 (out of a possible range of 0 to 1, will be accepted.