

# MORE DFT BENCHMARKS – SUGARS AND “MINDLESS” TEST SETS

The first is an examination by Csonka and French of the ability of DFT to predict the relative energy of carbohydrate conformation energies.<sup>1</sup> They examined 15 conformers of  $\alpha$ - and  $\beta$ -D-allopyranose, fifteen conformations of 3,6-anhydro-4-*O*-methyl-D-galactitol and four conformers of  $\beta$ -D-glucopyranose. The energies were referenced against those obtained at MP2/a-cc-pVTZ(-f)//B3LYP/6-31+G\*. (This unusual basis set lacks the *f* functions on heavy atoms and *d* and diffuse functions on H.) Among the many comparisons and conclusions are the following: B3LYP is not the best functional for the sugars, in fact all other tested hybrid functional did better, with MO5-2X giving the best results. They suggest the MO5-2X/6-311+G\*\*//MO5-2x/6-31+G\* is the preferred model for sugars, except for evaluating the difference between  ${}^1C_4$  and  ${}^4C_1$  conformers, where they opt for PBE/6-31+G\*\*.

This is really not a “mindless” study (as the term is used by Schaefer and Schleyer<sup>3</sup> and discussed in this post, where all searching is done in a totally automated way) but rather Grimme describes a procedure for removing biases in the test set. Selection of “artificial molecules” is made by first deciding how many

atoms are to be present and what will be the distribution of elements. In their two samples, they select systems having 8 atoms. The two sets differ by the distribution of the elements. The first set the atoms Na-Cl are one-third as probable as the elements Li-F, which are one-third as probable as H. The second set has the probability distribution similar to those found in naturally occurring organic compounds. The eight atoms, randomly selected by the computer, are placed in the corners of a cube and allowed to optimize (this is reminiscent of the “mindless” procedure of Schaefer and Schleyer<sup>3</sup>). This process generates a selection of random bonding environments along with open- and closed shell species, and removes (to a large degree) the biases of previous test sets, which are often skewed towards small molecules, ones where accurate experiments are available or geared towards a select group of molecules of interest. Energies are then computed using a variety of functional and compared to the energy at CCSD(T)/CBS. The bottom line is that the functional nicely group along the rungs defined by Perdew:<sup>4</sup> LDA is the poorest performer, GGA does much better, the third rung of meta-GGA functionals are slightly better than GGA functionals, hybrids do better still, and the fifth rung functionals (double hybrids) perform quite well. Also of interest is that CCSD(T)/cc-pVDZ gives quite large errors and so Grimme cautions against using this small basis set.

Source: <http://comporgchem.com/blog/?p=7>