Subject: Re: Bug in how stereochemistry is reported
Posted by thomas on Wed, 13 Jan 2021 16:51:04 GMT
View Forum Message <> Reply to Message

Hi Simon,

DataWarrior puts a lot of attention to handle stereochemistry correctly. When reading SD-files that have 3D coordinates, then the 3D-coordinates are taken to determine stereo parities. I tried to do the following to verify, whether there is a problem:
- I generated a file with 1000 random drug-like molecules
- I added a column with the stereo center count (found 0 to 5 stereo centers) and removed all rows with 0 stereo centers
- I added 3D-coordinates ending up with 480 molecules with 2D- and 3D-coordinates; none of them were racemic
- I added a column with row numbers
- I saved the file as SD-file (version 2) using 3D- rather than 2D-coordinates
- I merged the SD-file into the original file using the Row-No as merge key
- Then I checked, whether the structures were perceived as equal
- Except for 2 rows, molecules were identical. These two rows had unspecified double bond geometries in the first place and the conformer generator picked randomly one, which was different from the original. Thus, stereo centers were matching in all cases.

To understand what doesn't work in your case, I need to understand better, what you do. One reason may be that the software that you use to generate your 3D-coordinates and to export SD-files exports the molecules as racemates, i.e. with 'chiral flag' set to 0. Please check your SD-file with a text editor. The forth line (or every forth line after a $$$$ line) is called 'counts line'. It contains the number of atoms and bonds and other information about the molecule. It looks like:
 49 50 0 0 1 0 0 0 0 0999 V2000
If the fifth entry is a '1', then the structure is meant to be enantiomerically pure. If it is a '0', then this means that you have a racemic mixture containing the following structure plus its mirror image. If DataWarrior reads a racemate from a molfile, it shows green stereo bonds with a '&' sign that says, it is this and the opposite stereo configuration. Stereo bonds of absolute stereo centers are drawn in red. In case of racemates, DataWarrior does not necessarily display the enantiomer of the SD-file, because DataWarrior stores structures in a canonical way. For racemic mixtures it doesn't matter, which enantiomer is stored as long as it is marked as racemic. For exact molecules matching, however, it is important that also racemic mixtures are stored in a reproducible way.

If this is not the reason for you observation, please give me an example that doesn't work to find the reason.

Hope, I could explain in an understandable way...

Thomas