Subject: Loss of stereochemical information when importing sdf file
Posted by Mike Parker on Fri, 08 Feb 2019 20:11:31 GMT
View Forum Message <> Reply to Message

I have a database file in sdf file format. It includes chemical compounds with a single stereocenter in which stereochemistry is drawn explicitly. Compounds are either the (S)-enantiomer or the (R)-enantiomer. When I open the file in Datawarrior, each is depicted as a racemate. The stereochemical information appears to have been lost. However, when I copy the SMILES string for each structure I can see that a stereochemical designator is still present (@ or @@) - but the structures that were R and S in the original sdf file have both become S in the SMILES string. So, in fact, the stereochemical information appears to have actually been altered in some cases, not lost.

Any ideas about what might be going on or suggestions on how to avoid the issue?

Subject: Re: Loss of stereochemical information when importing sdf file
Posted by thomas on Tue, 12 Feb 2019 19:39:39 GMT
View Forum Message <> Reply to Message

I assume that your SD-File contains racemates. SD-files version 2 contain the so-called chiral-flag (5th entry in 4th line of every molfile), which defines, whether a molecule is a racemate or pure enantiomer. If this flag is 0, then even if the molecule contains up/down stereo bonds it must be interpreted as racemic.

After reading these SD-files DataWarrior knows that the stereo centers are racemic and not unknown (DataWarrior distinguishes this). Smiles do not distinguish between racemic and absolute stereo centers. Thus, if a racemic molecule has multiple stereo centers, DataWarrior writes the stereo configuration of these (@ and @@) into the smiles in order to retain the relative configuration. In the special case where the molecule only contains one stereo center, DataWarrior might omit the stereo information, because there is no relative stereo configuration. Currently, it does not, but I should probably change that.