Subject: Re: feature suggestion: .sdf export with explicit hydrogens
Posted by thomas on Thu, 23 Sep 2021 09:19:08 GMT
View Forum Message <> Reply to Message

Dear Norwid,

structures in DataWarrior files are stored as canonical encoded strings (idcodes). Therefore, it doesn't matter, whether a user draws a structure with or without hydrogen atoms. They are immediately converted into the canonical form, which is the one without hydrogen atoms. Tetrahedral stereo configurations are encoded as atom parity, such that when displaying the structure, one of the bonds will be drawn as up or down bond to account for the correct parity. This bond may be different from the one originally used to define the parity. As a consequence, when molecules are exported within SD-files, there are no simple hydrogens, but the stereo information is never lost. In order to display or export molecules with the originally supplied coordinates, DataWarrior stores coordinates of all idcode atoms as a separately encoded string. For 3D-coordinates there is a special handling: idcodes are still the same canonical ones without hydrogen atoms, but the coordinate encoding contains 3D-coordinates even for implicit hydrogen atoms, because for conformers hydrogen positions may be important, e.g. for docking. Therefore, SD-Files, if exported with 3D-coordinates, contain hydrogen atoms.

If I provided a setting to also write hydrogen atoms into exported 2D-SD-Files, then I would need to generate new potentially sub-optimal coordinates for the added formerly implicit hydrogen atoms. For 2D-SD-Files it seems common practice to not include implicit hydrogen atoms. Which application do you have in mind that requires hydrogen atoms?

Thomas