Subject: Re: feature suggestion: .sdf export with explicit hydrogens
Posted by nbehrnd on Sat, 25 Sep 2021 16:01:32 GMT
View Forum Message <> Reply to Message

Dear Thomas,

based on recent work with DataWarrior, I retract this feature suggest.

DataWarrior's default assignment of an idcode about a structure is a simplified, yet complete description of a molecule in terms of atom connectivity, as well as stereo chemistry. Thus -- already at the level where libraries of molecules are generated -- DW is able to assign e.g., a a full InChI string, despite at this stage, exported .sdf files lack lines with an explicit label "hydrogen". These .sdf aim for the exchange of information with other programs; including their own (independent from DW) assignment of full InChI strings.

The (re)generation of 3D structures by the conformer generator departing from such a "library .sdf" serves a different purpose. As you mentioned, these subsequent structures are the ones to be used in docking experiments. And these are the .sdf aiming for a visualization of the molecule's shape/geometry in a viewer like Jmol with explicit hydrogen atoms.

As noticed, the set of InChI strings assigned to molecules at level of library generation, and the one about structures past the conformer generation, may differ from each other. This however is because the conformer generator resolves ambiguity if an entry in a "library stage .sdf" did not describe a stereogenic center in either (R), or (S); an axis in (P), or (M); a double bond in (E), or (Z) configuration. The conformer generator's permutation of these choices may be caused by lack of this particular information, i.e. an unknown stereo configuration, or because the .sdf read intentionally describes a racemate (the "enhanced stereo recognition" the structure editor's documentation describes).

Norwid