Subject: translation of the idcodes
Posted by nbehrnd on Thu, 22 Oct 2020 18:05:24 GMT
View Forum Message <> Reply to Message

When choosing to save (some of) the content a .dwar file into the .txt file, the column about the structures contains the idcode. This condensed string suffices to recreate the molecular structures when this .txt file is read in a pristine session of DataWarrior. Contrasting to, e.g. a SMILES string, it is harder to understand «the grammar» behind this efficient storage of structural information.

Possibly this is less a feature suggest for DW, rather than bridging DW to other programs: Are there plans to to render the idcode format accessible, either as reading input into for, or (/and?) equally as written output by OpenBabel, e.g. for one of the sections 18.1 till and including 18.3 of the current manual about formats in cheminformatics?(1)

Norwid

(1) https://open-babel.readthedocs.io/_/downloads/en/latest/pdf/

Subject: Re: translation of the idcodes
Posted by thomas on Fri, 23 Oct 2020 13:05:22 GMT
View Forum Message <> Reply to Message

The idcode is a canonical very compact text representation of a molecule.
Its creation requires ring and aromaticity perception and much more complicated, complete enhanced stereo perception. It also includes sub-structure query features and this way is somewhat comparable to an isomeric SMARTS. The best way for another software package to create or read it, is to use the original source code, which is available for anybody and any purpose in the OpenChemLib project on Github.

LigandScout is an example of a commercial software package that can read idcodes and DataWarrior conformer files.

Whether OpenBabel will support OpenChemLib idcodes once, is not in our hands.

Thomas