Subject: C=C double bonds/search by extended stereochemistry
Posted by nbehrnd on Tue, 30 Nov 2021 16:11:02 GMT
View Forum Message <> Reply to Message

Hello Thomas,

my understanding of DataWarrior's manual is that the program internally stores the molecules in a format capable to discern and (if selecting .sdf v3000) export of double bonds including those without explicit assignment of (E)/(Z) configuration. As anticipated, the SMILES strings

C/C=C/C
C/C=C\C
CC=CC

yield three different representations of but-2-ene.

For a small test set (cf. double_bond.dwar and the screen photos in the .zip archive attached), I aimed to use the features of extended stereochemisty as a criterion to retain a sub set (Chemisty -> From Chemical Structure -> Add Substructure Count), including the unassigned double bond visualized as a crossed double bond. So far, the attempts failed.

As anticipated, the search pattern of (cis)-butene, retained both (cis)-butene and cyclohexene and withdrew the oxime. However -- contrasting to anticipation -- (trans)-butene and the unassigned butene equally were retained. Does the search for, e.g. crossed double bonds, require a particular instruction?

Version info: DataWarrior 5.5.0 for Linux in Debian 12/bookworm (branch testing), including the update as «development version» as fetched on Sunday, 2021-11-28.

Norwid

File Attachments

1) double_bonds.dwar, downloaded 557 times
2) double_bond_documentation.zip, downloaded 554 times

Subject: Re: C=C double bonds/search by extended stereochemistry
Posted by thomas on Tue, 11 Jan 2022 15:12:52 GMT
View Forum Message <> Reply to Message

Dear Norwid,

sorry for the long delay. I somehow missed your question. As default DataWarrior does not match the stereo configuration of atoms or bonds on the substructure search. You may, however, require for individual atoms or bonds a stereo configuration match. In the editor select the lasso tool and

double click an atom or bond. In the dialog that opens, you can select 'Match Stereo Configuration'. Then DataWarrior uses the 'Enhanced Stereo Representation' rules defined by MDL many years ago, e.g. a racemic center in the query matches a racemic, OR, and pure R- or S-center. An explicit R in the query only matches a pure R. For stereo bonds you cannot select 'Match Stereo' for cross bonds, which is in-line with the atom-logic, because an unknown configuration always matches all better specified configurations in the MDL logic. This means you can directly search for E- and Z-bonds, but not for unspecified (cross). You could, however, do that with two individual inverted filters, i.e. excluding E- and excluding Z-.