Subject: Bug in how stereochemistry is reported
Posted by richards99 on Fri, 08 Jan 2021 10:52:09 GMT
View Forum Message <> Reply to Message

Hi,
When a molecule has had its 3D coordinates generated in other software packages, and then these are loaded into DataWarrior as an SDF, it seems the stereochemistry representation in the table is not always correct or defined.

Not knowing how this works so I am only speculating, but I suspect DW may be picking up flags just from a 2D representation and not looking at the 3D representation. Is it possible for DW to look and see if 3D coordinates have been generated, and if they have, then to display the stereochemistry defined in the 3D conformation.

It has been causing a lot of problems, in that when generating smiles in DW of molecules with 3D coordinates, it is sometimes returning the wrong stereochemistry or undefined stereochemistry.

Thanks,

Simon.

Subject: Re: Bug in how stereochemistry is reported
Posted by thomas on Wed, 13 Jan 2021 16:51:04 GMT
View Forum Message <> Reply to Message

Hi Simon,

DataWarrior puts a lot of attention to handle stereochemistry correctly. When reading SD-files that have 3D coordinates, then the 3D-coordinates are taken to determine stereo parities. I tried to do the following to verify, whether there is a problem:
- I generated a file with 1000 random drug-like molecules
- I added a column with the stereo center count (found 0 to 5 stereo centers) and removed all rows with 0 stereo centers
- I added 3D-coordinates ending up with 480 molecules with 2D- and 3D-coordinates; none of them were racemic
- I added a column with row numbers
- I saved the file as SD-file (version 2) using 3D- rather than 2D-coordinates
- I merged the SD-file into the original file using the Row-No as merge key
- Then I checked, whether the structures were perceived as equal
- Except for 2 rows, molecules were identical. These two rows had unspecified double bond geometries in the first place and the conformer generator picked randomly one, which was different from the original. Thus, stereo centers were matching in all cases.

To understand what doesn't work in your case, I need to understand better, what you do. One reason may be that the software that you use to generate your 3D-coordinates and to export SD-files exports the molecules as racemates, i.e. with 'chiral flag' set to 0. Please check your SD-file with a text editor. The forth line (or every forth line after a $$$$ line) is called 'counts line'.

It contains the number of atoms and bonds and other information about the molecule. It looks like:
 49 50 0 0 1 0 0 0 0 0999 V2000
If the fifth entry is a '1', then the structure is meant to be enantiomerically pure. If it is a '0', then this means that you have a racemic mixture containing the following structure plus its mirror image. If DataWarrior reads a racemate from a molfile, it shows green stereo bonds with a '&' sign that says, it is this and the opposite stereo configuration. Stereo bonds of absolute stereo centers are drawn in red. In case of racemates, DataWarrior does not necessarily display the enantiomer of the SD-file, because DataWarrior stores structures in a canonical way. For racemic mixtures it doesn't matter, which enantiomer is stored as long as it is marked as racemic. For exact molecules matching, however, it is important that also racemic mixtures are stored in a reproducible way.

If this is not the reason for you observation, please give me an example that doesn't work to find the reason.

Hope, I could explain in an understandable way...

Thomas

Subject: Re: Bug in how stereochemistry is reported
Posted by richards99 on Thu, 04 Mar 2021 08:44:58 GMT
View Forum Message <> Reply to Message

Thanks Thomas,
Looking into it, it does appear to be related as to whether the structure is specified as chiral or not.

The problem is various tools I am using are not compatible with V3000 flags, and simply writes out structures SDF format with the appropriate R or S stereochemistry and nothing else, so no '1' as the fifth entry for instance.
Then when I read into DataWarrior, and then save back out to an SDF, the stereochemistry can often be flipped to what it was originally. Presumably because DW doesn't care which isomer it shows as it thinks there is a mixture anyway.
Even if DW believes the stereo centre is a mixture, it would still be useful if it did not alter the provided stereochemistry it was given.

Simon.

Subject: Re: Bug in how stereochemistry is reported
Posted by richards99 on Thu, 04 Mar 2021 08:50:12 GMT
View Forum Message <> Reply to Message

There also seems to be issues with it accepting chiral atoms next to Nitrile groups. For example, if you paste this made up structure into DataWarrior, it should have two chiral atoms, as defined in the SMILES, but it does not create two chiral atoms:

CC[C@H](O)[C@@H](CC1=CC=CC=C1)C#N

Thanks,
Simon.

Subject: Re: Bug in how stereochemistry is reported
Posted by thomas on Tue, 09 Mar 2021 15:01:52 GMT
View Forum Message <> Reply to Message

Dear Simon,

I tried to confirm the issue with CC[C@H](O)[C@@H](CC1=CC=CC=C1)C#N with the official 5.2.1 version and with the current dev version. If I paste the smiles into an editor or if I paste a tiny table with the smiles being part of it, DataWarrior always created the correct structure with two stereo centers. Can you please give me a procedure where it doesn't work?

The issue with the chiral flag not set in the SD-file is a V2000 problem, not V3000. The fifth entry in the counts line of a V2000 molfile encodes, whether the structure is a racemate or not (0 or 1). For racemates DataWarrior must normalise the configuration, because DataWarrior stores canonical structure representations. A normalization means here that it determines in a reproducible way, which of the two enantiomers is shown and stored. It cannot just store the input form, because then there would be two different representations encoding the same structure.

What I could do, is to provide an option when loading V2000 SD-files to consider racemic molfiles with encoded stereo centers as enantiomerically pure, just assuming that the chiral flag is meant to be set. This, of course, is dangerious, especially if more than one stereo centers are present. Depending on the source of the molfile, however, one may have good reason to believe that the software 'forgot' to set the chiral flag, e.g. when the structures represent conformers with 3D-coordinates.

What do you think?

Thomas

Subject: Re: Bug in how stereochemistry is reported
Posted by richards99 on Thu, 11 Mar 2021 08:26:49 GMT
View Forum Message <> Reply to Message

Hi Thomas,
Below is a text file of the SMILES I mention in the previous post. When I load this directly into DataWarrior, the carbon with the alcohol attached correctly shows a wedge bond indicating its chirality, whilst the carbon which has the nitrile attached only shows flat bonds with a ? label to indicate unknown stereochemistry. I am using the Dev build of DataWarrior. Other colleagues can reproduce the problem I am seeing also.

With regards the chiral flag issues of mol files, the solution you propose would be very useful as there seems to be multiple tools which are not setting the chiral flag, and therefore when importing into DataWarrior the stereochemistry is getting scrambled. Whether it is in the dialog box for opening an SDF, or a setting in the preferences, I do not mind, so long as an option is available.

Thanks,

Simon.

File Attachments
1) temp.txt, downloaded 819 times

Subject: Re: Bug in how stereochemistry is reported
Posted by richards99 on Thu, 11 Mar 2021 12:19:41 GMT
View Forum Message <> Reply to Message

Have just downloaded the latest DEV update last night, and the Nitrile groups are now showing with a chiral wedge bond. So whatever changes have been done, it is fixed in the new DEV build.

Thank you.

Simon.

Subject: Re: Bug in how stereochemistry is reported
Posted by thomas on Fri, 12 Mar 2021 14:34:22 GMT
View Forum Message <> Reply to Message

I just updated the dev version such that when loading an sd-file, then it interprets the content:
If all entries are V2000 and if zero entries have the chiral flag set and if some entries have stereo centers,
then the user is asked, whether entries with stereo centers shall be interpreted as pure enantiomers.
The macro task 'Open File' also got an option to assume the chiral flag to be set for such cases.

Subject: Re: Bug in how stereochemistry is reported
Posted by richards99 on Mon, 15 Mar 2021 08:21:18 GMT
View Forum Message <> Reply to Message

Brilliant, thanks Thomas. This is going to save a lot of problems we have been suffering from!

Simon.

Subject: Re: Bug in how stereochemistry is reported
Posted by tryckmans@ridgelinediscov on Mon, 24 Jul 2023 12:01:44 GMT
View Forum Message <> Reply to Message

Hello Thomas,
I have an issue similar to the one above. When generating a sdf file using ChemDoodle, when I open the sdf in Datawarrior the wrong stereoisomer is created. I notice the SDF does not contain a chiral flag as described above (the 5th character on line 4 is "0", not "1".) When contacted, ChemDoodle answered me this
"While I cannot help with DataWarrior, they are likely handling the SDF file in un-inverted y-axis coordinates. In 2D computer applications, the y-axis is inverted, so the reading application needs to take this into account when loading the SDF file if it contains a 2D drawing. Every chemical drawing application inverts the y-axis in this way."

many thanks for your comments!

The SMILE gives the correct (S) configuration:
N[C@@H](CC(=CC=C1)C=C1)C(O)=O

SD file below:
Molecule Name
  ChemDodl07242313512D 0   0.00000   0.00000   0
[Insert Comment Here]
  0 0 0   0 0   999 V3000
M V30 BEGIN CTAB
M V30 COUNTS 12 12 0 0 0
M V30 BEGIN ATOM
M V30 1 N 0.8511 1.0249 -1.0000 0
M V30 2 C 0.8803 0.0253 -1.0000 0
M V30 3 C 1.7604 -0.4492 1.0000 0
M V30 4 C 0.0292 -0.4997 -1.0000 0
M V30 5 O 2.6115 0.0759 1.0000 0
M V30 6 O 1.7896 -1.4489 1.0000 0
M V30 7 C -0.8511 -0.0252 -1.0000 0
M V30 8 C -0.8803 0.9744 -1.0000 0
M V30 9 C -1.7022 -0.5503 -1.0000 0
M V30 10 C -1.7604 1.4489 -1.0000 0
M V30 11 C -2.5823 -0.0758 -1.0000 0
M V30 12 C -2.6115 0.9238 -1.0000 0
M V30 END ATOM
M V30 BEGIN BOND
M V30 1 1 1 2
M V30 2 1 2 4
M V30 3 1 4 7
M V30 4 1 7 8
M V30 5 2 7 9
M V30 6 1 9 11
M V30 7 2 11 12
M V30 8 1 12 10

M V30 9 2 10 8
M V30 10 1 2 3 CFG=3
M V30 11 1 3 6
M V30 12 2 3 5
M V30 END BOND
M V30 BEGIN COLLECTION
M V30 MDLV30/STEABS ATOMS=(1 2)
M V30 END COLLECTION
M V30 END CTAB
M END
> <DATE>
24-07-2023

$$$$

Subject: Re: Bug in how stereochemistry is reported
Posted by thomas on Thu, 27 Jul 2023 17:45:01 GMT
View Forum Message <> Reply to Message

Dear Tom,

molfiles V3000 don't use a 'chiral' flag. If you open your molfile in DataWarrior, you will notice that it shows an absolute stereo center. Of course, DataWarrior uses the correct Y-direction for molfiles, which is indeed inverted in comparison to most other xy-orientations. CFG=3 is a down bond and coordinates are correctly created leading to an R-isomer. ChemDraw does the same. I am afraid that ChemDoodle created the wrong CFG entry here.

Best wishes,

Thomas
