Subject: Aromaticity perception
Posted by richards99 on Thu, 25 Jun 2020 20:38:01 GMT
View Forum Message <> Reply to Message

Hi,
There appears to be consistent issues with the import of certain aromatised structures.
Ones I especially notice are N-methylated pyridinones, pyrimidinones, and bicyclic structures.
Data Warrior puts double bonds into the wrong positions, quaternising nitrogens.

It would be useful if the aromaticity perception can be improved otherwise it creates lots of invalid smiles which require manually altering.

Thanks,

Simon.

Subject: Re: Aromaticity perception
Posted by nbehrnd on Fri, 26 Jun 2020 08:13:52 GMT
View Forum Message <> Reply to Message

Hi Simon,

may you add from which software / file format you import into DW to observe these issues and share a typical minimal input/output file? For now, it reads related to the (intermediate) representation of 2-pyridone with SMILES either as lactam C1=CC=CNC(=O)1, or lactim Oc1ccccn1.

Norwid

Subject: Re: Aromaticity perception
Posted by richards99 on Sat, 27 Jun 2020 22:09:10 GMT
View Forum Message <> Reply to Message

Hi Norwid,
I cannot find a way to attach an SDF file, but basically this set of smiles below causes a problem.
If you create a CSV/TXT file of these and import them into DW, then all is fine.
But if you create an SDF file of these smiles using Marvin Sketch for example (kept as aromatised), and then import them into DW, then all the double bonds are messed up!
Other programmes such as Marvin, RDKit do not seem to have issues with these.

Cn1ccccc1=O
Cn1cccnc1=O
Cn1ccncc1=O
Cn1ccn(C)c(=O)c1=O
Cn1ccc(=O)cc1

Cn1ccc2ncccc2c1=O
Cn1ccc(=O)c2cccnc12
Cn1c2cccnc2ccc1=O

Subject: Re: Aromaticity perception
Posted by richards99 on Sat, 27 Jun 2020 22:14:27 GMT
View Forum Message <> Reply to Message

Okay, worked out now how to attach the file.
Attached is the SDF file which is problematic for DW, and a snapshot showing the difference between importing as smiles or SDF.

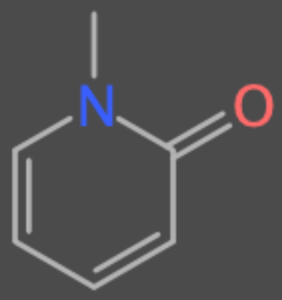
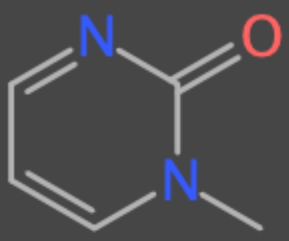
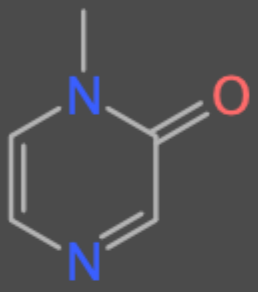
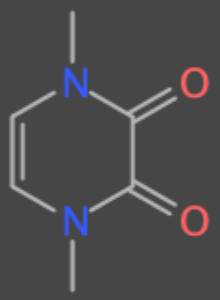
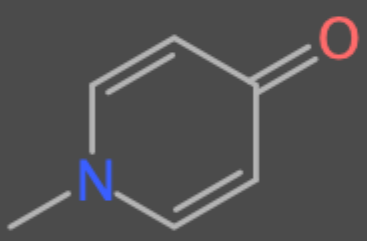
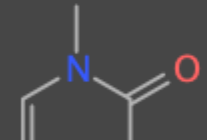
Simon.

File Attachments

1) Screenshot 2020-06-27 at 23.06.40.png, downloaded 1332 times

smiles.csv

Table

	Structure of ...	Molecule
1		Cn1ccccc1=O
2		Cn1cccnc1=O
3		Cn1ccncc1=O
4		Cn1ccn(C)c(=O)c1=O
5		Cn1ccc(=O)cc1
6		Cn1ccc2ncccc2c1=O

Table

2) Aromatised.sdf, downloaded 738 times

Subject: Re: Aromaticity perception
Posted by nbehrnd on Sun, 28 Jun 2020 21:57:00 GMT
View Forum Message <> Reply to Message

Hi Simon,

it was possible to replicate the problem with the .sdf shared by you. Tentatively, the problem is caused by the existence of multiple SMILES dialects by different programs which may be an obstacle for DW. I thus recommend to pass the .sdf to openbabel to pass the files' content into a new .sdf file to solve the issue. Here, DataWarrior (version 5.2.1, native installation in Linux Debian) and openbabel (version 3.1.0, June 9, 2020) were used.

Enter the directory containing the .sdf in question. From the terminal (Linux, Mac) or cmd.exe (Windows) provide a instruction in line of

obabel -isdf Aromatised.sdf -osdf -O Aromatised_passed_obabel.sdf

With only eight molecules, this is a quick operation. Comparing the original and the new / derived .sdf file with each other shows that the connection table in the files is adjusted, as shown in the screen photo below:

More importantly, the issues with tetravalent nitrogen atoms are resolved:

The .dwar eventually obtained is provided as an attachment of this answer.

Openbabel is a freely available program running on Windows, Mac, and Linux to interconvert chemical formats. Its code is open on GitHub, which equally hosts the executables. The documentation may be accessed online, or offline. If wanted, a GUI may provide you an easier entry into a selection of its functions, too.

It is equally possible to convert the SMILES as provided into an .sdf, too. The command then would be

obabel -ismi probe.smi -osdf -O probe.sdf

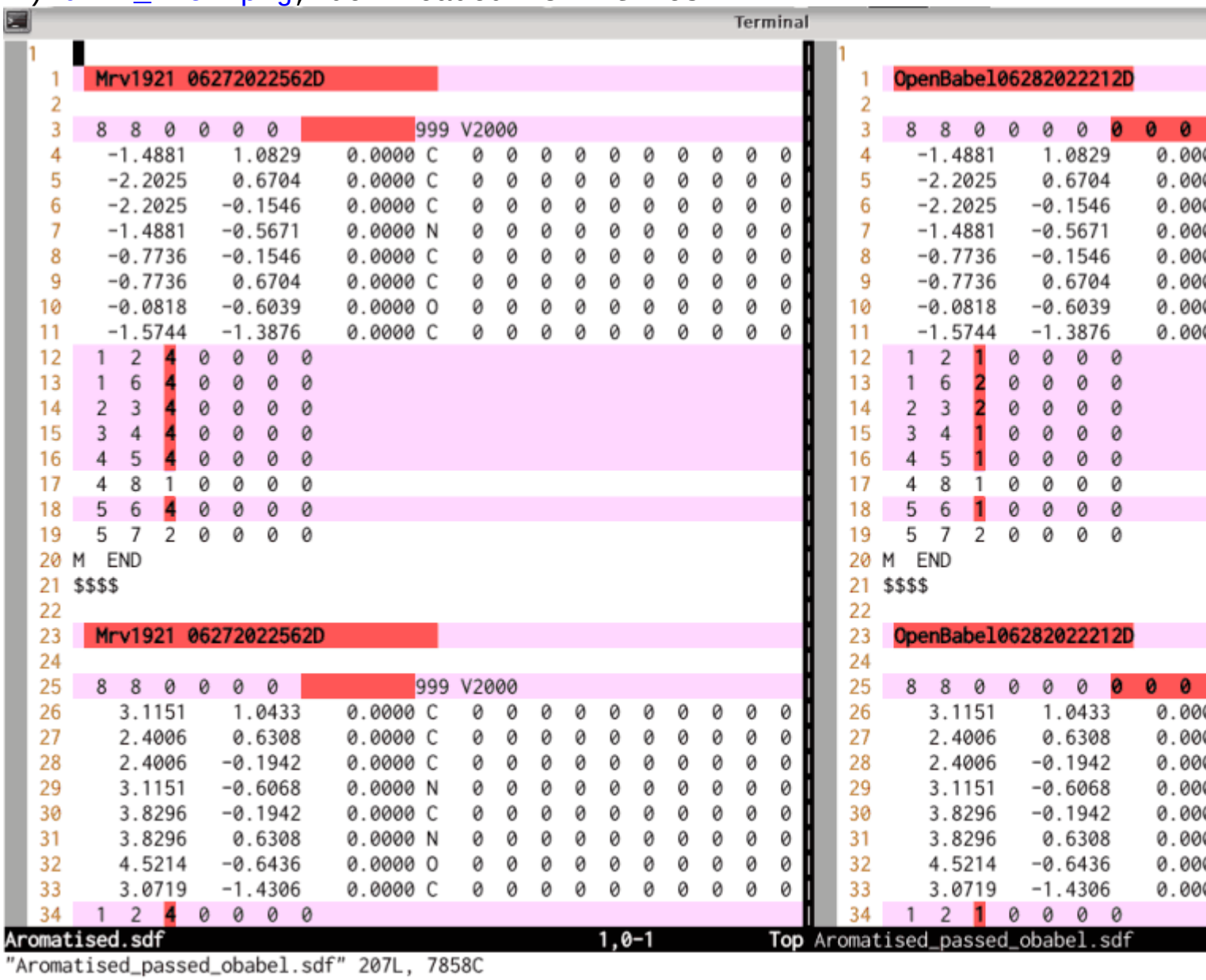
to lead to the same result as above, or copy-pasting (without header row) the SMILES directly into DW. Both .smi and .sdf of this approach equally are provided here.

Norwid

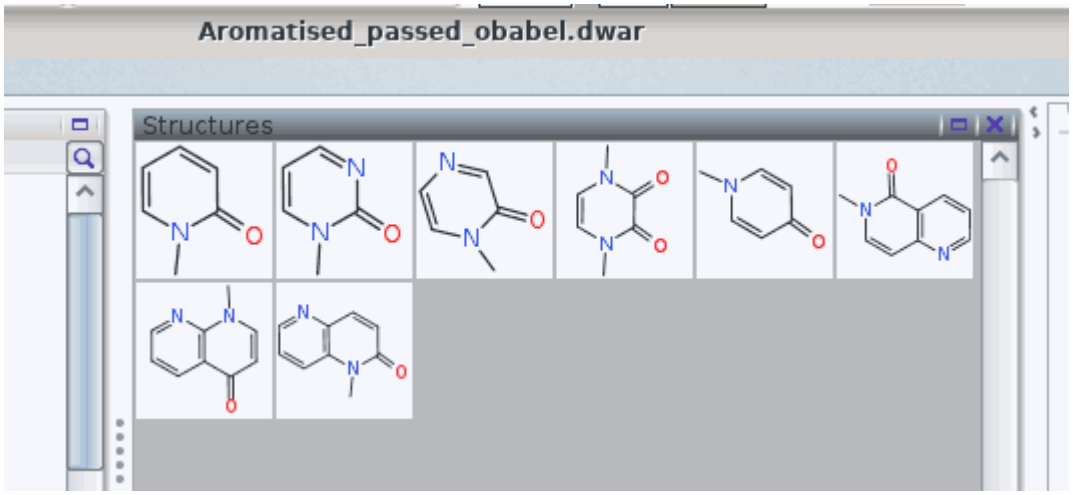
https://github.com/openbabel/openbabel
https://github.com/openbabel/openbabel/releases/tag/openbabe l-3-1-1
https://open-babel.readthedocs.io/en/latest/
https://open-babel.readthedocs.io/_/downloads/en/latest/pdf/

File Attachments

1) diff_view.png, downloaded 1327 times



2) Aromatised_passed_obabel-or8.png, downloaded 1353 times



3) Aromatised_passed_obabel.dwar, downloaded 747 times
4) probe.sdf, downloaded 730 times
5) probe.smi, downloaded 799 times

Subject: Re: Aromaticity perception
Posted by thomas on Thu, 02 Jul 2020 12:41:03 GMT
View Forum Message <> Reply to Message

many thanks Simon and Norwid for pointing to this issue suggesting work-arounds.

The problem was that DataWarrior didn't expect finding compounds with aromatic bond types in molfiles,
which are based on the Daylight aromaticity model into the bargain, e.g. having carbonyl carbon atoms being
maked as aromatic. This is unusual for two reasons: First, molfiles typically store alternating single and double bonds for aromatic rings rather than using the delocalized bond type, unless it encodes
a substructure with query features. Second, for the rare cases that the delocalized bond types may be used
one would expect an MDL/Symex/Hueckel aromaticity concept to be applied.

Nevertheless, since Marvin Sketch seems to read SMILES based atom aromaticity encodings and writes them directly into
written molfiles using aromatic bond types, I have updated DataWarrior to normalize this kind of encoding before
generating and writing idcodes (DataWarrior's canonical structure representation) into its native files.

The current developments version should not have this issue anymore.

Thomas