

---

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.

```
Cn1cccc1=O
Cn1ccnc1=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 1414 times

Table

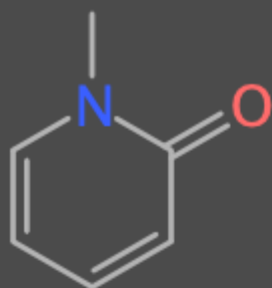


Structure of ...

Molecule

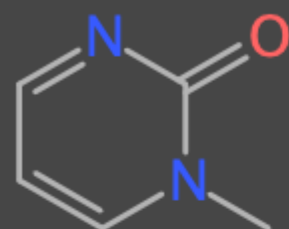


1

Cn1cccc1=O

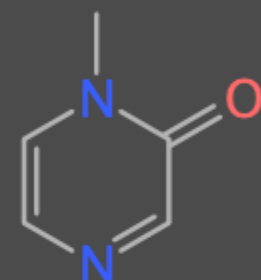
1

2

Cn1cccnc1=O

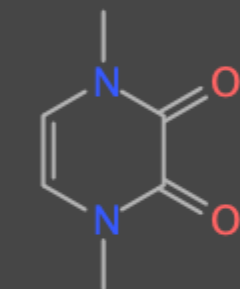
2

3

Cn1ccncc1=O

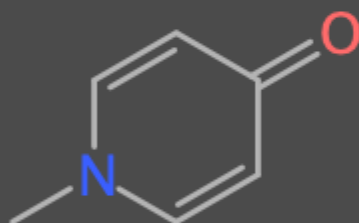
3

4

Cn1ccn(C)c(=O)c1=O

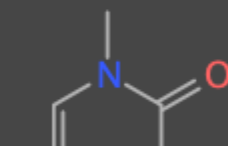
4

5

Cn1ccc(=O)cc1

5

6

Cn1ccc2ncccc2c1=O

6

2) [Aromatised.sdf](#), downloaded 769 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/openbabel-3-1-1>

<https://open-babel.readthedocs.io/en/latest/>

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

## File Attachments

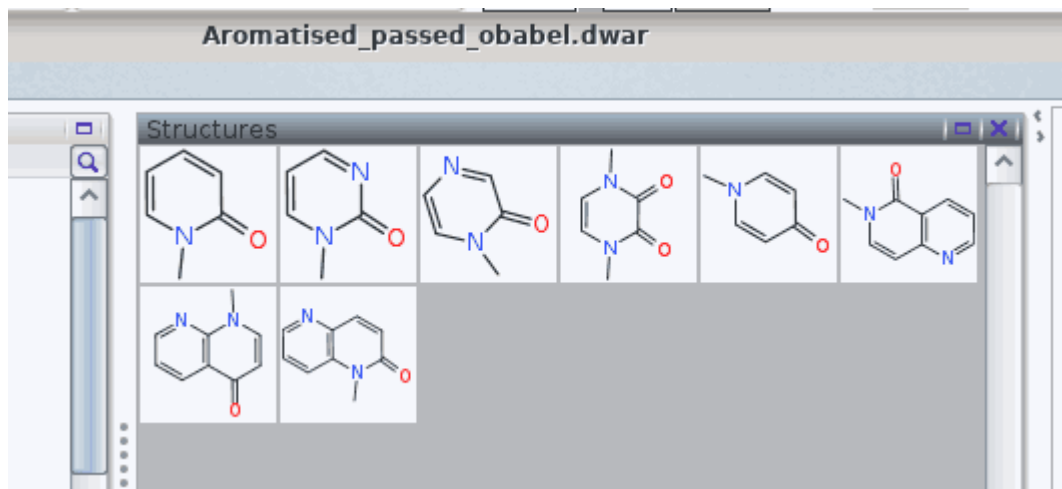
1) [diff\\_view.png](#), downloaded 1418 times

```
Terminal
1 Mrv1921_06272022562D
2
3 8 8 0 0 0 0 999 V2000
4 -1.4881 1.0829 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5 -2.2025 0.6704 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
6 -2.2025 -0.1546 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
7 -1.4881 -0.5671 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
8 -0.7736 -0.1546 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
9 -0.7736 0.6704 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
10 -0.0818 -0.6039 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
11 -1.5744 -1.3876 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
12 1 2 4 0 0 0 0
13 1 6 4 0 0 0 0
14 2 3 4 0 0 0 0
15 3 4 4 0 0 0 0
16 4 5 4 0 0 0 0
17 4 8 1 0 0 0 0
18 5 6 4 0 0 0 0
19 5 7 2 0 0 0 0
20 M END
21 $$$$
22
23 Mrv1921_06272022562D
24
25 8 8 0 0 0 0 999 V2000
26 3.1151 1.0433 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
27 2.4006 0.6308 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
28 2.4006 -0.1942 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
29 3.1151 -0.6068 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
30 3.8296 -0.1942 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
31 3.8296 0.6308 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
32 4.5214 -0.6436 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
33 3.0719 -1.4306 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
34 1 2 4 0 0 0 0

1 OpenBabel06282022212D
2
3 8 8 0 0 0 0 0 0 0 0
4 -1.4881 1.0829 0.0000
5 -2.2025 0.6704 0.0000
6 -2.2025 -0.1546 0.0000
7 -1.4881 -0.5671 0.0000
8 -0.7736 -0.1546 0.0000
9 -0.7736 0.6704 0.0000
10 -0.0818 -0.6039 0.0000
11 -1.5744 -1.3876 0.0000
12 1 2 1 0 0 0 0
13 1 6 2 0 0 0 0
14 2 3 2 0 0 0 0
15 3 4 1 0 0 0 0
16 4 5 1 0 0 0 0
17 4 8 1 0 0 0 0
18 5 6 1 0 0 0 0
19 5 7 2 0 0 0 0
20 M END
21 $$$$
22
23 OpenBabel06282022212D
24
25 8 8 0 0 0 0 0 0 0 0
26 3.1151 1.0433 0.0000
27 2.4006 0.6308 0.0000
28 2.4006 -0.1942 0.0000
29 3.1151 -0.6068 0.0000
30 3.8296 -0.1942 0.0000
31 3.8296 0.6308 0.0000
32 4.5214 -0.6436 0.0000
33 3.0719 -1.4306 0.0000
34 1 2 1 0 0 0 0

Aromatised.sdf 1,0-1 Top Aromatised_passed_obabel.sdf
"Aromatised_passed_obabel.sdf" 207L, 7858C
```

2) [Aromatised\\_passed\\_obabel-or8.png](#), downloaded 1438 times



- 3) [Aromatised\\_passed\\_obabel.dwar](#), downloaded 773 times
- 4) [probe.sdf](#), downloaded 752 times
- 5) [probe.smi](#), downloaded 824 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 made 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

---