
Subject: Suggestion: tautomer-check

Posted by [nbehrnd](#) on Wed, 06 May 2020 13:40:09 GMT

[View Forum Message](#) <> [Reply to Message](#)

DataWarrior's model to assign Druglikeness depends on the encoded structure a tautomer is represented. With *Ambit-Tautomer*, Kochev et al. published an open source tool (Java based), to predict tautomers and to rank their likelihood.

If wanted, their executable may be run without the larger *Ambit* framework, just by

```
java -jar ambit-tautomers-2.0.0-SNAPSHOT.jar
```

Thus, I would like to suggest DataWarrior could implement a function to check if the structures to consider could reasonably yield a tautomer worth to probe equally. This equally could be complementary to your recently published reference `tautomer.dwar`.

Out of curiosity, I drew a pyridone, a pyrazole, and a thalomid with ACD ChemSketch in two tautomeric forms, exported the SMILES strings (as defined by ACD ChemSketch) into a `.smi` file:

With `openbabel`, it was converted into a `.sdf` accessible for DataWarrior by

```
obabel -ismi tautomers.smi -osd -O tautomers.sdf
```

successfully read and used to compute the Druglikeness. At least as the examples about «hydroxypyridine» and the enol form of the thalomid differ in the results examined.

To ease replication of the findings, the relevant files are provided below.

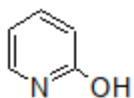
Publication about *Ambit-Tautomer*: <https://doi.org/10.1002/minf.201200133>

github-entry about *Ambit-Tautomer*:

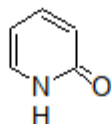
<https://github.com/ideaconsult/apps-ambit/tree/master/tautomers-example>

File Attachments

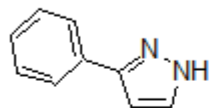
1) [tautomers.gif](#), downloaded 1043 times



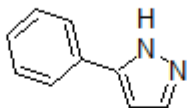
pyridin-2-ol
Oc1ccccn1



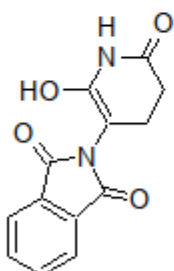
pyridin-2(1H)-one
O=C1C=CC=CN1



3-phenyl-1H-pyrazole
c1ccc(cc1)c2ccnn2

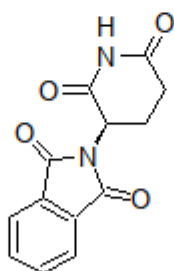


5-phenyl-1H-pyrazole
c1ccc(cc1)c2ccnn2



thalomid_enone

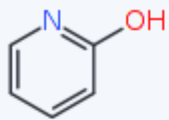
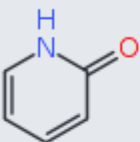
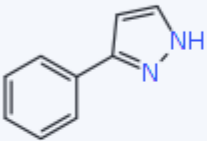
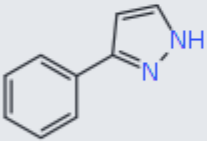
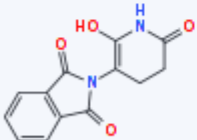
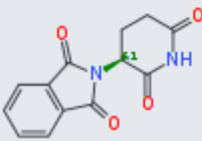
OC=1NC(=O)CCC=1N3C(=O)c2ccccc2C3=O



thalomid_ketone

O=C1NC(=O)CC[C@H]1N3C(=O)c2ccccc2C3=O

- 2) [tautomers.smi](#), downloaded 695 times
- 3) [tautomers.sdf](#), downloaded 681 times
- 4) [tautomers.dwar](#), downloaded 706 times
- 5) [tautomer_DW.png](#), downloaded 1180 times

	Structure	Molecule Name	H-Acceptors	H-Donors	Druglikeness
1		pyridin-2-ol	2	1	-1.88
2		pyridin-2(1H)-one	2	1	-0.23382
3		3-phenyl-1H-pyrazole	2	1	1.3013
4		5-phenyl-1H-pyrazole	2	1	1.3013
5		thalomid_enone	6	2	1.7671
6		thalomid_ketone	6	1	4.3067

Subject: Re: Suggestion: tautomer-check

Posted by [thomas](#) on Mon, 11 May 2020 18:59:48 GMT

[View Forum Message](#) <> [Reply to Message](#)

This is, of course, an important issue. Having the right tautomer to start with is crucial for any reliable prediction.

The intention behind the work on the published tautomer database is to allow others and ourselves to work on algorithms to predict the most prominent tautomer. If we had such an algorithm with a reasonable performance at hand, we could use it as input before training any prediction model. It also would allow to normalize pKa databases to improve or develop algorithms to more reliably predict pKa-values, which is also in big demand. We have it in the pipeline, but it will take time...