I notice that sometimes DW will display a given tautomer/protomer different than what was written to the SDF file.

For example, I have the canonical smiles:

Nc1nc([O-])c2[nH]c[nH+]c2n1

which DW will display as:

Nc(nc1O)nc2c1[nH]cn2

I seem to be able to prevent this by adding explicit Hs when writing the SDF. I'd just like to have a clearer picture of what it is doing.

thanks!

DataWarrior stores chemical structures in memory and within its native files as very compact canonical string representation, so-called idcodes. SD-Files and SMILES are converted to idcodes, when they are read as input file (or from the clipboard). After reading the original, DataWarrior performs some structure normalization before creating the canonical idcode. One example is to unify the nitro group as -N(+)(=O)-O(-), which in some input SMILES or SD-files comes as -N(=O)=O. Part of the normalization is also to neutralize charged atoms as your example:

Because of these normalization rules DataWarrior considers these both structures to be equivalent and, thus, creates the same idcode for them, which actually represents the second form of the structure.

Note that you can manually change this structure again within DataWarrior by adding charges and implicitly adding/removing hydrogen atoms to come back to the first structure. If you save the file and reopen, the bipolar structure is retained. The idea behind that is that structure files from different sources may use different conventions for representing the same thing. DataWarrior does its best to normalize. When a chemist, however, directly within DataWarrior edits a structure, then he probably means exactly what he is drawing.

## File Attachments

1) t.png, downloaded 884 times