Subject: Re: TFA salts
Posted by thomas on Mon, 15 Jun 2020 20:07:01 GMT
View Forum Message <> Reply to Message

This is indeed a difficult question. First, when DataWarrior converts SMILES into its native format (called idcodes), it run this three step process:
- convert the SMILES into a chemical structure
- normalize the structure, e.g. convert non-polar-5-valent nitro to the polar 4-valent nitro, or distribute protons to remove charges
- then encode the structures in a canonical way after uniquely assigning atom numbers

The second step is responsible that the triflate is stored as mixture if two non polar compounds. The normalization and canonicalization are done for DataWarrior to recognize the same compounds as being the same, even if they are drawn in a different way. Admittedly, the stored form is a formal one and may not always reflect the reality, e.g. in case of neutralized zwitter ions or salts. However, if you manually change the protonation state, then DataWarrior accepts your wish to overwrite and keeps it permanently.

When calculating properties using an algorithm with a predefined list of fragments, then it makes a difference, whether we use the salt or neutral form. The cLogP and TPSA algorithm both consider charges, thus the bipolar structures are likely to produce better results.

For DataWarrior there is a dilemma. Shall it trust the input protonation state or should it predict proper protonation depending on predicted pKas (which it cannot do, currently)? By normalizing it assures that the same structures are always recognized as being the same and also treated the same.

Hope this explains it, even if it is not fully satisfactory.

Thomas