Subject: Errors calculating flexophore-based similarity/activity cliff of certain molecules
Posted by user on Tue, 11 May 2021 07:35:12 GMT
View Forum Message <> Reply to Message

Flexophore-based similarity/activity cliff analysis of a dataset that contains non-standard SMILES of tetrazoles no longer works after upgrading to datawarrior 5.5.0 (used to work in datawarrior 5.2.1).

Error message I got:

Standardized SMILES of the same compounds didn't cause any errors in datawarrior 5.5.0.

When importing the SMILES strings from a CSV file it seems that the SMILES to structure conversion also gives weird bonds, which may cause this issue (manually pasting the SMILES into structure editor gives no errors). Other compounds may also have this issues since I got like hundards of error messages while working on a diverse dataset of over 7K molecules, but I have only confirmed on tetrazoles.

An example .csv file and the resulting problematic .dwar file that could reproduce the problem is attached for reference.

File Attachments
1) error.csv, downloaded 501 times
2) error.dwar, downloaded 473 times
3) error.png, downloaded 766 times

Subject: Re: Errors calculating flexophore-based similarity/activity cliff of certain molecules
Posted by thomas on Tue, 11 May 2021 15:35:58 GMT
View Forum Message <> Reply to Message

Thank you for sending these strange, but correctly defined molecule examples. The charge normalization did indeed make a mistake with these 1,2-hydrogen shifted 1,2 dipolar input structures. The problem in the charge normalization is fixed and another issue with the flexophore not working in case of one pharmacophore point only is fixed as well. You can download the update as development patch from the DataWarrior download page after clicking the 'read and understood' checkbox. Note, that the download link for this dev update is in the fine print.

Subject: Re: Errors calculating flexophore-based similarity/activity cliff of certain molecules
Posted by user on Thu, 13 May 2021 04:39:49 GMT
View Forum Message <> Reply to Message

Edit: the windows version of dev patch works now, but not for the linux version.

Thanks for the quick response Thomas! Flexophore analysis of the dataset is now working with the dev patch (for windows).

Those molecule examples were part of the smiles generated by an open source de novo fragment-based virtual screening code that I'm currently playing with. That code has a feature of automatically generating multiple tautomers and protonation states of a compound, I guess that's where those weirdly defined structures come from.

Datawarrior make it much easier to visualize the screening results with minimal scripting, and the 2D structure of smiles in datawarrior looks nicer than the figures generated with rdkit especially for those bulky and more complicated structures. Really appreciate all your hard work developing this software and making it publicly available.