Subject: Re: Drug Likeness vs Drug Score
Posted by thomas on Tue, 02 Feb 2021 22:27:52 GMT
View Forum Message <> Reply to Message

The druglikeness is based on a long list of (>5000) substructures of which every one is associated with a numerical value. These substructures basically list all more or less frequent substructures to be found in drug and/or commercially available building blocks. If a fragment is a frequent substructure in building blocks, but rarely found within drugs, then it get a negative value that quantifies this ratio. Fragments that are predominantly found in drug structures have positive values. For calculating the druglikeness, DataWarrior goes through the list of substructures. When a substructure is part of the query molecules, then its value is added to the sum. Finally, the sum is normalized by deviding it by SQRT(fragmentCount).
Building block like molecules get values below zero, while drug like molecules get positive sum. A bar chart in the DataWarrior manual (above link) shows that.

If you are interested to look at the fragments with contributions, you may download the data file from the DataWarrior source code:

https://github.com/Actelion/openchemlib/blob/master/src/main
/resources/resources/druglikenessNoIndex.txt

Open it in a text editor, add a first line 'idcode value' (without quotes; make sure that the white space after 'idcode' is a TAB), and save the file. Then you can open it in DataWarrior, where you should see the structures and associated contribution values.

Thomas