Subject: Re: Macro for evolutionary library (EL) in Datwarrior (DW)
Posted by thomas on Sun, 11 Sep 2022 13:33:37 GMT
View Forum Message <> Reply to Message

Dear Norwid,

many thanks for answering all the questions already...

Dear Jon,

let me quickly add some comments to two questions:

If you want to keep running the newest version: Once install the original 5.5.0 version from 2021 (which you have already done). Once in a which download the dw550win.zip (Mac or Linux dw550x.zip) using the small print link from the official datawarrior download page. Unpack the file to obtain a few replacement files. Make a save copy of the respective original files with the same names and move the files from the zip archive to the DataWarrior installation folder, where they replace the original files. To move to an earlier version you can just copy an earlier set of files or the saved original files back to the DataWarrior installation folder.

I am aware that the explanation of how to use the evolutionary library and how it works is not well represented in the manual. Especially, the new fitness criteria PheSA and docking have a lot of potential. The best thing to understand what it really does one might look a the source code, but admittedly, that is not for everybody and it still takes some time even for a person being fluent in Java. I also realized that the current algorithm could be improved by reusing fragments of previous hi-ranking structures. The current principle is rather simple: apply simple modifications to the best structures of the previous generation, rank them and use the best ranking molecules for further modifications. Modifications are selected randomly from a complete list of available modifications. The random selection is completely random: modifications are the more likely, the more the product is drug-like (or natural product like). This ensures that created molecules make sense in this regard. The central Java class that performs the changes is called Mutator and can be found here. It lists the kind of changes possible and also contains the few lines of code to actually perform the change in the molecule. The rest is basically how to calculate fitness, which is a weighted sum of user selected parameters. The more complex are Flexophore similarity (DOI:10.1021/ci700359j), PheSA (paper will come) and Docking.

Hope this is somewhat useful...

Thomas