
Subject: similarity_search_protocol

Posted by [khom](#) on Mon, 20 Sep 2021 20:09:07 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear DW forum,

My goal is to query how similar a small (fewer than 10) set of molecules are to those in a big (about 50,000) pool. I am wondering if the following protocol is sound and efficient. 1) From the chemical structures of the molecules in the big set, calculate the descriptors: FragFp; PathSp; SphereFp; SkelSphere; OrgFunctions; Flexophore, generate similarity chart and neighbor tree for each descriptor and save as file_a. 2) Open the small query list of compounds, select the "Find Similar compounds in file" button, specify file_a as the target file, and save similar compounds found to a new file. Or should I add the new queries to the existing big file, and re-calculate the similarity scores for the entire big(ger) file? Thank you in advance for your inputs.

Subject: Re: similarity_search_protocol

Posted by [thomas](#) on Thu, 23 Sep 2021 10:10:29 GMT

[View Forum Message](#) <> [Reply to Message](#)

The fastest way would be to open the large file, then add a Structure-List (similarity) filter and within the filter with a right mouse click read the query structures into the filter. The play with the similarity slider. SkeletonSpheres would represent chemical similarity best. You can use Flexophores for potential similarity of target activity.

You could calculate conformers for both, the big and the query files. Then, open the big file and do a conformer superpositioning with the query compounds. This would give you shape and pharmacophore point similarity as well as a visual feedback of the best matches.

You could append the big with the query file and create a SOM with maybe 150 nodes per dimension. This gives you an chemical space visualization with where your query molecules are in regard to the 50000.

And as suggested, the 'Find similar compounds in file' calculates all similarities and gives you all the pairs of similar compounds.
