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.