Subject: Similarity Search
Posted by Rishi on Mon, 28 Oct 2024 13:55:40 GMT
View Forum Message <> Reply to Message

Hi,

I want to conduct a similarity search for one compound against ~60K compounds within the same set. I'm only really interested in the similarity between the one specific compound against the 60K set, and not the similarity between all other compounds within the set.

When I run a search such as the FragFP similarity search the estimated time is >50 hours. I assume this is because it's calculating similarity between all compounds with the 60K set.

Is there a way of just conducting the search on the one compound against the set to reduce calculation times of similarity based positions?

Many thanks!

Subject: Re: Similarity Search
Posted by thomas on Mon, 28 Oct 2024 15:39:24 GMT
View Forum Message <> Reply to Message

If you have an open DataWarrior window with 60k compounds and draw a structure in the structure filter to perform a similarity search, the search should be immediate. What exactly do you do?

Subject: Re: Similarity Search
Posted by Rishi on Mon, 28 Oct 2024 15:50:52 GMT
View Forum Message <> Reply to Message

I want to conduct both FragFP searches and also flexophore searches, but on one specific compound against the larger set

Subject: Re: Similarity Search
Posted by thomas on Thu, 31 Oct 2024 09:18:57 GMT
View Forum Message <> Reply to Message

That should be straight-forward:
- open the file with the 60k compounds
- if you don't see a structure filter, then add a new one from Edit->New Filter
- choose similarity(FragFp) or (Flexophore) in the filter and draw or paste in the query struccture
(if option "similarity (flexophore)" is missing, you need to calculate flexophores (Chemistry->From Chemical Structure->Calculate Descriptor->Flexophore"

FragFp filtering should be immediate, flexophore filtering may take a few seconds. If the filtering doesn't seem to finish, then the filter process crashed. Then, most likely there is something wrong with the data. You can try then to delete the FragFp or Flexophore column and recreate it. Then try again. Or send me the file (ideally just a few rows that have the same problem) and I check...