Subject: Re: Similarity analysis using "find similar compounds..." - slow analysis of libraries
Posted by SM2020 on Fri, 27 Nov 2020 00:35:09 GMT
View Forum Message <> Reply to Message

Thanks very much Thomas!! - that will be a really useful update. Looking forward to that rolling out soon.

Getting a distribution of molecules within a single file would be great, but comparing 2 discrete files - and plotting a resulting similarity distribution - is more what I was thinking about. In my mind they are slightly different (but I could be mistaken).

Maybe I'm misunderstanding, but using your suggestion, could I just put all the cpds into a single file, run the analysis you suggest and then (assuming there is a unique identifier for the cpds - e.g. vendor), use DW to highlight by vendor and then assess the distribution in the resulting histogram? Would this compound selection step be possible (like in other graphical displays in DW)?

I was essentially thinking to keep one library constant (reference) and compare against "other" libraries thereby generating a binned histogram of closet neighbour similarities for each "other" library compared to the reference - one measure of similarity between "other" libraries and a reference library is then gauged by the apparent shift of distributions between (near) 0 and 1.

Either way, I'd like to try both analyses. It's great that DW (will) allow it.

Is there a hard upper limit for the no. of compounds analysed in these ways (i.e. file size, > 16K cpds?? or is it driven by available computational resources to carry out the task?

Thanks again for your effort.