Subject: Methods to select diverse compounds using non-binary descriptors?
Posted by EJW515 on Wed, 30 Jun 2021 11:37:04 GMT
View Forum Message <> Reply to Message

Hi all,

I have a large library of compounds (100k) from which I'm trying to select a diverse set. The "select diverse compounds" function has worked for FragFp and Spheres fingerprints.

What may be too big a computational task for larger libraries is to select compounds based on conformational diversity too. For this, the Flexophore descriptor worked very nicely on a smaller set of 2000 molecules. I clustered the set into 30 groups (which took 3 minutes), from which the 30 "representative" molecules could serve as a diverse subset. To me, clustering with a non-binary descriptor appears to be analogous to using the "select diverse set" command with a binary descriptor (FragFp, SpheresFp).

However, trying to cluster a library of 100k into 180 groups by Flexophore gave me a waiting time of 500 hours... My setup is a Dell Precision 7920/Intel Xeon Gold 5122 4-core/NVIDIA Quadro P1000/64 GB RAM.

Has anyone here ever tried to do something similar?

I wondered if there's a better approach in some manual way, but the calculation time just for "maxsim(Flexophore)" on the set of 2000 already took 1.5 hours. I have a feeling this may be a fundamental limitation due to the complexity of calculating Flexophore similarity (relative to Tanimoto coefficients), but it would be a very useful tool for me so I thought it worth asking.

Thanks for any help :)

Best,
EJW515