Subject: Re: AllFragFp
Posted by thomas on Wed, 24 Apr 2024 12:45:18 GMT
View Forum Message <> Reply to Message

AllFragFp is substantially different from FragFp: it is hashed and uses 2048 bits. AllFragFp internally generates all substructures of a given molecule with up to 6 connected bonds including stereo chemistry. These substructures are converted into a canonical representation from which a hash code between 0 to 2047 is generated, for which the corresponding bit is set. The original idea was to accelerate the substructure search by a more discriminating descriptor than the FragFp. If the AllFragFp descriptor is available in a DataWarrior file, then DataWarrior uses that for substructure pre-screening. Since the sub-structure search is usually fast for mot more than some hundred thousand molecules, one shouldn't bother to use the AllFragFp. For many millions, however, it makes a significant difference.

Regarding the value of similarities calculated by this descriptor, I didn't really investigate it applicability domain. It certainly will produce very fine grained similarity values, but the SkeletonSpheres descriptor is will probably generate more intuitive ones, because by design single atom replacements cause less large losses of similarity compared to other substructure based descriptors.