Subject: Globularity
Posted by P_Fitz on Wed, 24 Mar 2021 01:53:23 GMT
View Forum Message <> Reply to Message

I'm dabbling a bit in chemoinformatics and was wondering how a globularity score (specifically as referred to by eNTRy rules https://www.nature.com/articles/nature22308.pdf) compares to the Shape Index that is calculated in data warrior.

Thanks so much for developing and supporting such a great tool for free. I've found it very useful in my research.

Patrick Fitzgerald

Subject: Re: Globularity
Posted by thomas on Wed, 24 Mar 2021 16:13:09 GMT
View Forum Message <> Reply to Message

Hi Patrick,

the shape index is based on the 2D-graph of the molecule and looks at the those two atoms that have the longest distance in terms of bonds between them. This length and the atom count of the molecule are used to calculate the shape index.

The globularity is calculated from atoms coordinates in conformers. It is basically doing a singular value decomposition to rotate the molecule in the coordinate system such that its width (size in x-direction) is maximized. Then it is rotated around the x-axis until its height is maximized. After that globularity is calculated as the molecule size in z-direction devided by the size in x-direction.

Shape index and globularity have a slight negative correlation, but one can hardly used as substitute for the other. I have just added a globularity calculation to DataWarrior (dev version). The results correlate (0.83) with the Nature paper's glob value. This is not a perfect correlation. I use up to 128 non-energy minimized conformers and average the result. If I minimize conformers with MMFF94s+ or reduce their number and therefore overpopulate low energy conformers, then the correlation gets worse. Probably, the conformers used in the paper were random conformers with no bias on low energy and most likely not forcefield minimized.

Hope this helps,

Thomas

Subject: Re: Globularity
Posted by thomas on Fri, 26 Mar 2021 16:13:16 GMT
View Forum Message <> Reply to Message

Hi Patrick again,

I just deployed an update, because the recent version was based on the singular values and created too small globularity values. Strangely, the Nature dataset's globularity values also seem too small. Now my calculation uses a singular value decomposition to determine the three orthogonal axes through the molecule that represent largest, second largest and smallest variance of the atom coordinates. Then, I determine along the axes the size of the molecule and divide the smallest by the largest size resulting in a reasonable globularity value. I average individual values of up to 32 random conformers, which were not minimized, but produced with a bias for low energy. The correlation with the 'glob' values of the Nature dataset is smaller than before, but if you explore conformers, create surfaces and visually inspect the molecules, it makes much more sense. One more thing to mention is that the conformers contain all hydrogen atoms, which gives on average slightly larger values than without hydrogens.

Thomas

Subject: Re: Globularity
Posted by P_Fitz on Fri, 26 Mar 2021 23:31:14 GMT
View Forum Message <> Reply to Message

Thank you so much for the detailed replies. I'm excited to use the new functionality!

Subject: Re: Globularity
Posted by thomas on Tue, 30 Mar 2021 15:10:58 GMT
View Forum Message <> Reply to Message

I made another update with another method to calculate globularity based on the volume and area of the solvent excluded surface (Van der Waals surface). The globularity is the surface of a perfect sphere, whose volume exactly matches the molecule's volume, divided by the real surface of the molecule. In my opinion this is the most reasonable globularity calculation. I also added some documentation and explanation.

Thomas
