

6.3: Discussion

While there are many molecular fingerprints and similarity coefficients, it is not feasible to use all possible combination of them for a given project with limited time and resources. For this reason there have been many studies that compared performances among different fingerprints and similarity coefficients. In their large-scale analysis of 37 molecular descriptors [1], Bender and coworkers evaluated similarity between the descriptors and identified four broad descriptor classes: (1) circular fingerprints, (2) circular fingerprints considering counts, (3) path-based fingerprints and structural keys, and (4) pharmacophoric descriptors. This study suggests that the performance of the descriptors is much more defined by those four classes than the particular parametrization used or individual descriptors. This implies that descriptors that belong to the same class are likely to give similar results (e.g., similar hit compound lists) when they are used for molecular similarity evaluation.

In general, the Tanimoto coefficient is a preferred metric for molecular similarity comparison, but Dice and Cosine coefficients are considered as good alternatives [2]. For example, a study by Bajusz and Héberger [2] compared eight well-known similarity distance metrics on a large data set of molecular fingerprints. This study concluded that the Tanimoto, Dice, Cosine, and Soergel coefficients are the best metrics for similarity calculation, in the sense that they produce the most similar rankings to those averaged over the rankings produced by the eight similarity metrics considered. The Euclidean and Manhattan distances were found to be not optimal because they gave different rankings from other metrics.

Further Reading

- Molecular Similarity in Medicinal Chemistry
<https://doi.org/10.1021/jm401411z>
- Molecular similarity: a key technique in molecular informatics
<https://doi.org/10.1039/B409813G>
- Daylight Theory: Fingerprints
<https://www.daylight.com/dayhtml/doc/theory/theory.finger.html>
- How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space
<https://doi.org/10.1021/ci800249s>
- Extended-Connectivity Fingerprints
<https://doi.org/10.1021/ci100050t>

References

1. Bender A, Jenkins JL, Scheiber J, Sukuru SCK, Glick M, Davies JW: **How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space.** J Chem Inf Model 2009, 49:108-119.
2. Bajusz D, Racz A, Héberger K: **Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?** J Cheminformatics 2015, 7:20.

6.3: Discussion is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.