

FTrees - Search fast, bridge scaffolds

Christine Dramburg, **Ingo Dramburg**, Christian Lemmen
BioSolve IT GmbH
An der Ziegelei 75
53757 Sankt Augustin, Germany
dramburg@biosolveit.de

Feature Trees is a fast similarity searching method, that utilizes a tree representation for a descriptor[1]. It is independent of the three-dimensional structure of the compounds, but takes their topology of chemophysical properties into account. Briem and Lessel compared previously seven different similarity descriptors including FTrees in an affinity fingerprint experiment over five activity classes[2]. In their comparison of mean hit rates the Feature Trees descriptor outperformed the others.

Recently Bender et. al. used the same dataset for an evaluation of atom environment descriptor [3]. In their approach they used a Naive Bayesian approach for classification. We compared all available searching approaches currently implemented in the FTrees software, match-search, split-search and the dynamic-search approach. Additionally, we used the MTrees approach which is capable of building models by combining several Feature Trees into a single one. The results have been analyzed according to multiple performance measures and compared with the published results.

We demonstrate that Feature-Trees compares favourably to other approaches. Not all search methods perform equally well. The results vary significantly for different input datasets. However, on average FTrees outperforms the other approaches. Especially interesting is the combination of multiple query molecules. Hereby the number of retrieved actives can be significantly increased compared to a single molecule query. As it is not uncommon to have several known actives at hand, this should be of practical relevance for a lot of similarity-based search methods.

In addition to classical virtual screening, FTrees can be used for de-novo-design based on its Fragment Space extension[5]. FTrees-FS is able to generate similar compounds via a fragment buildup process based on a set of fragments and so-called linker-rules. It has been shown that the approach is able to generate reasonable hits from fragment spaces of theoretical sizes of up to 10^{18} drug-size compounds within a few minutes of computer time. The selection of different search methods and the adjustable similarity range of the retrieved results allows for finding promising compounds that are structurally dissimilar but exhibit similar properties. Here we show results on a Fragment Space which we recently generated based on publicly available compounds.

References

- [1] Rarey M, Dixon JS, *Journal of Computer Aided Molecular Design*, 1998, **12**, 471.
- [2] Briem H, Lessel UF, *Perspectives in Drug Discovery and Design*, 2000, **20**, 231.
- [3] Bender A, Mussa HY, Glen RC, *Journal of Chemical Information and Computational Science*, 2004, **44**, 170.

[4] Zimmermann M, PhD thesis at the university of Bonn, **2004**.

[5] Rarey M, Stahl M, *Journal of Computer Aided Molecular Design*, 2001, **15**,497 .