JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMPUTER SOFTWARE REVIEW

pubs.acs.org/JACS

Software Review of FTrees and FTrees-FS in Pipeline Pilot

FTrees and FTrees-FS in Pipeline Pilot. BioSolveIT GmbH. An der Zieglei 79, 53757 Sankt Augustin, Germany. http://www.biosolveit.de/FTrees. See Web site for pricing information.

FTrees is a similarity search tool for querying large databases of compounds represented by their Feature Tree descriptor, a tree structure where functional groups and rings are reduced to nodes (see Rarey, M.; Dixon, J. S. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471−490). The concept of FTrees dates back to work done by Matthias Rarey in 1997 during a six-month research stay at SmithKlineBeecham (USA). Since 2002, the FTrees software has been marketed and further developed by BioSolveIT. FTrees-FS (Feature Trees Fragment Spaces) is an extension of the original algorithm. It allows searching of large virtual combinatorial libraries without the need for enumeration, only storing building blocks and connection rules (see Rarey, M.; Stahl, M. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 497−520).

Both FTrees and its extension module, FTrees-FS, have been implemented in Pipeline Pilot (PP). For this review, FTrees version 2.3.0 was used as part of the FTrees PP interface package 2.3.0.1 under Pipeline Pilot version 8.0.1. The package comprises four components: FTrees Calculator—to generate the FTrees descriptor; FTrees Similarity—to calculate similarities between molecules and queries; and an FTrees Reader and Writer—to manage calculated FTrees. The extension module FTrees-FS is available as a separate interface. It comprises the FTrees-FS component, which generates hit molecules for a given query from a supplied Fragment Space file. All calculations were performed on a single processor.

General Comments. It is worth mentioning that no constraints are set as to what a user is allowed to do with a demo license and that there is no feature restriction either. There initially were some technical issues with the PP components provided and with the license management, i.e., locating a license or the number of licenses that were needed to run the software, as each separate component requires a license. However, the BioSolveIT team was very responsive and eager to help and debug components. The FTrees in PP user guides are useful to get set up relatively quickly, albeit additional information regarding parameter settings for individual components would have been helpful. Specific documentation about FTrees can be found in the extensive user-guide manual for the command line version. Various expert features, which would be desirable for special applications, are not customizable via the Pipeline Pilot interface, but the features provided are sufficient for routine analog searching. Initial test runs are important even for users who are familiar with the FTrees command line version. In particular, it is advisible to inspect the FTrees log file to ensure FTrees are produced with correct donor and acceptor features as anticipated. FTrees produced on the command line can be visualized with FlexV; however, the PP version generates FTrees in a special format that cannot be viewed with this tool and cannot be visualized in PP. On the other hand, a useful feature that has been introduced in PP is the visualization of query and hit molecule mappings in the resulting Excel file. These maps show matching substructures (in identical colors) in both molecules along with their local (substructure) similarity scores and an overall FTrees similarity score.

FTrees. To investigate the performance of FTrees, five kinase inhibitors were chosen as queries, each corresponding to a different chemotype. In-house assay data were available with thousands of known actives. Virtual screening was undertaken against our corporate screening collection comprising a few million compounds including the known actives. The performance of FTrees was compared to other typical ligand-based screening methods: Topological Pharmacophores (Schneider et al. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896), Reduced Graphs (Harper et al. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2145−2156), and Daylight Fingerprints (Daylight Toolkit, Daylight Chemical Information Systems Inc., Laguna Niguel, CA). It was determined whether the query was able to find itself and analogs (trivial) or to identify other known chemotypes, i.e., assess the scope for scaffold hopping. Furthermore, whether the methods would find any interesting new chemotypes for testing was assessed, as no High Throughput Screen had been performed on this target. The corporate library was converted overnight to FTrees. This was followed by 2−3 h for similarity calculation without map generation for each of the five queries with an FTree similarity threshold set to 0.7/0.8 to speed up processing time. In contrast, the search time for the other methods using a single processor and a similar threshold was just seconds to minutes.

All methods were able to find their starting point and analogs and to scaffold hop to some extent, as two queries were able to retrieve each other's chemotype in the top 200. However, all methods also missed other known chemotypes. They also found new chemotypes, which are subject to activity testing. FTrees appears to be more similar to the performance of Topological Pharmacophores than to that of Reduced Graphs in this particular case, but a conclusion cannot be drawn on this single observation and would require more extensive analysis on different sets of targets. FTrees seems to be a useful tool as part of an array of drug discovery tools and somewhat orthogonal to other methods. However, the running time of the Pipeline Pilot implementation takes much longer than all other methods investigated, but this can be addressed by using the parallel processing option in PP.

FTrees-FS. FTrees-FS has already seen real life applications in industry where bespoke Fragment Space for FTrees-FS was generated based on combinatorial libraries using BioSolveIT's CoLibri tool (Lessel et al. *J. Chem. Inf. Model.* **2009**, *49*, 270−279; Boehm, M. et al. *J. Med. Chem.* **2008**, *51*, 2468−2480). KnowledgeSpace is a ready-to-use, freely available Fragment Space that comes along with the FTrees-FS PP component and comprises ∼11K unique fragments that can be combined to produce 10^{10} synthetically accessible

Published: September 28, 2011

products. It is based on 82 published synthesis protocols from the *Journal of Combinatorial Chemistry*, which cover GPCRs, proteases, kinases, and chemistry-driven libraries. Here, the issue of whether analogs could be found within the supplied KnowledgeSpace was tested using the same five kinase inhibitor queries as before. Somewhat similar compounds with scores of 0.9 were found for certain queries, even in libraries not specifically designed for kinases. These can be useful for generating ideas for modifications. Interestingly, only one query had hits from a kinase library. Other hits with lower scores (<0.9) only had limited resemblance to the query. All hit structures seemed overly elaborate as starting points for a chemistry program.

Next, a trivial validation experiment was performed to see if a KnowledgeSpace hit finds itself at top rank when reused as a query, and it did. Finally, KnowledgeSpace was tested to see how it compares to a real-life Fragment Space. This question was addressed with a known TGFβ inhibitor previously used by Lessel et al. to search their own combinatorial chemistry Fragment Space. At least one hit with an FTrees score of 0.93 was quite similar to the query but featured a different core. However, the coverage of drug-like space would need to be tested more thoroughly using a more comprehensive set of known drug queries.

In conclusion, FTrees-FS is fast to run, taking a minute or less to return the top 100 hits along with their maps, and the free KnowledgeSpace may be useful for generating ideas, but the real value comes with a proprietary Fragment Space.

Monika Rella
GlaxoSmithKline

10.1021/ja208498e