

Welcome to *FTreesXL*

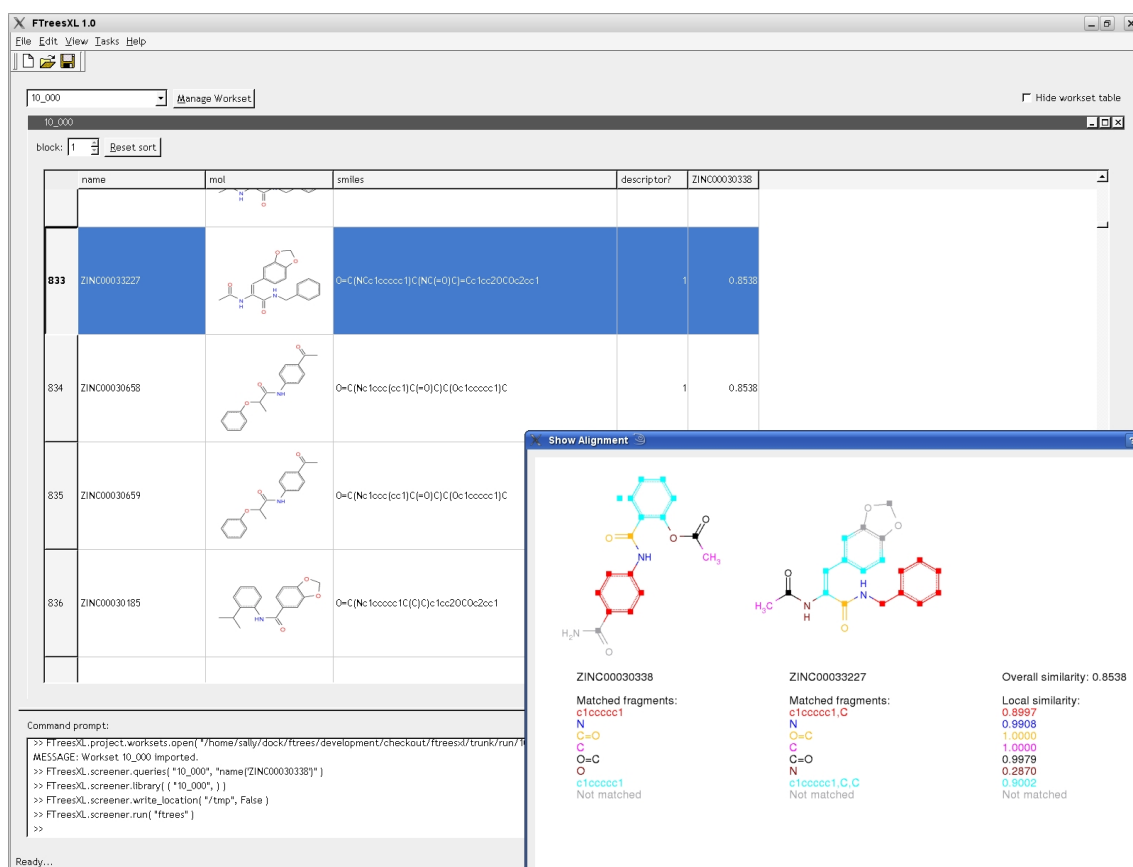


Figure 1.1: A typical view of *FTreesXL*

1.1 Tips Before You Get Started

1.1.1 Worksets in *FTreesXL*

The first step when working in *FTreesXL* is to begin a new workset. The workset is the basic "document" type used by *FTreesXL*. Worksets can be created and manipulated with the standard commands **New**, **Open**, **Save**, **Save As** and **Close**. The workset itself contains a summary of the following: the set of molecules with which you have been working, the Feature Trees (FTrees) you

created for these molecules and any other information that has been appended to the molecules, such as screening scores. Worksets are shown in the table. You may have several worksets open at once in *FTreesXL*, but only one workset may be viewed at once.

1.1.2 Molecule names

The linkage between molecules, *FTrees* and other associated information is based on the fact that the molecules have a set of unique identifiers. It was decided that the *molecule name* would form the unique identifier. It is, therefore, important that SMILES input is annotated with names and while MOL2 molecule names are fixed, you can choose the name of molecules from SD files as being either the first header line or a specific property field. So BEWARE: It is essential that molecule names in *FTreesXL* are unique or you may see undefined behaviour!

1.2 First Steps

The first step when working in *FTreesXL* is to begin a new workset (**File New**). You will see an empty table with the **Workset Manager** window open ready for you to start importing molecules. Import molecules by clicking **Select Molecules** and choosing files in the File Browser – choose a file and click **Open**; you may select several files, finally close the Browser with **Cancel**. A simple progress monitor is shown in the shell. Imported molecules are shown in the table. Next, move to the **Descriptors** tab in the **Workset Manager**. This is a very simple dialog that allows you to **Create** (or recreate, for example, if you change the parameters by **Edit Preferences**) and **Delete** *FTree* descriptors for the current workset. After creating descriptors, **Save** your workset. You are now ready to run calculations. Note: You can close the **Workset Manager** dialog and reopen it any time using the **Manage Workset** button. The **Workset Manager** always opens for the workset that is currently active.

1.3 Screening

The simplest way to carry out a screening calculation is to click a molecule in the table (it will be highlighted blue) and then open the **Screening** task dialog from the **Tasks** menu. The query is already chosen as being the molecule **currently highlighted in table**, the workset name should appear as the screening workset name for the screening **Library**. Now, simply click **Screen** and the calculation will start. At the end of the screening calculation you will see the similarity scores appended in a new column in the table. The name of this column can be changed and a comment for this column can be added using the context menu (right mouse click) on the column title. The context menu shown when clicking on the similarity score for a molecule brings up a window showing how the molecules align and how the similarity score is achieved.

1.4 Workset Manipulation

1.4.1 Selecting

The most useful facility is the **Selection** Dialog found in the **Select** tab of the **Workset Manager**. You will find a ? button in the section **Make New Selection** - this will show a small help dialog for making selections in the workset. Using the syntax shown in the help dialog, enter an expression in the space to make a new selection. The selection will be shown in the table according to the radio buttons in the **View Mode** section. In the command prompt section below the table you will see a message telling you how many molecules were selected. It is also possible to save selections

in the workset so that you can view them again/use them elsewhere in the tool. Give the selection a name in the space under **Remember new selection...** and click **Save** in Workset. This name is now available in several dropdown menus in this dialog and also in the Task dialogs.

1.4.2 Appending your own Data

Data contained in external files can be appended to the molecules using the molecule names as a matching key. As default the data must be in a space separated file with the format:

```
ID           <data col name>  <data col name>
<mol name>   <data1>          <data2>
etc.
```

In the **Data** tab use the **Import Data File** facility to select the data file and **Import** it to the workset. (There are other templates available for importing files created by the *FTrees* software as data.)

1.5 Clustering

To cluster, open the **Clustering** task dialog from the **Tasks** menu. There is a limit to the number of molecules you can cluster - set by default to 25 000. This may already take days to compute!! If you have a large workset it will be sensible to first make a selection on this workset and cluster just the selection. It is also possible to reuse an already computed matrix - it is the matrix calculation that takes up most of the time. The matrix must fit the molecule selection exactly (give the matrix a corresponding name so it will be easy to identify again!). There are several clustering algorithms available: choose one and use the **Edit Prefs** button to refine the parameter settings if necessary. Now, simply click **Cluster** and the calculation will start. At the end of the clustering calculation you will see the cluster IDs appended in a new column in the table. Note: cluster IDs are currently classed as properties, which are separate from screens. This is important for selection.

1.6 Exporting information

You are most likely to want to be able to export information from *FTreesXL* for further use elsewhere, or save more useful subsets of molecules. There are three possibilities:

1.6.1 Export selection to new workset

This can be found in the **Select** tab of the **Workset Manager**: enter the name of the selection and a new workset filename.

1.6.2 Export selection to MOL2 file

This can be found in the **Select** tab of the **Workset Manager**: enter the name of the selection and a new MOL2 filename.

1.6.3 Export data

This can be found in the **Data** tab of the **Workset Manager**: choose the data to export and enter a filename.

1.7 Logs and Scripting

You can see in the Command Prompt area below the table that many of the actions you carry out translate directly to a command. At any time you may enter a command here instead of clicking in the GUI. All the commands you see are saved at the end of each *FTreesXL* session in a log file according to the naming scheme:

```
/tmp/.FTreesXLlogs/FTreesXL_YYYY_mm_dd_hh_mm_ss
```

These files are removed when they reach seven days old. These files can be used again to run common workflows in one simple step. Copy the file to some location where it will be safe and edit it as appropriate - we will now call it a script. Scripts are loaded using **File Load Script...** When the script has finished the table will pop up and *FTreesXL* will be ready to continue from that point.

More information about *FTreesXL*

2.0.1 *FTreesXL* and *FTrees*

FTreesXL is a graphical user interface (GUI) to our highly efficient molecule similarity software *FTrees*. *FTrees*' primary application is in fuzzy similarity searching to facilitate virtual HTS. Due to the unique way its underlying topological descriptor (the Feature Tree) captures connectivity and physico-chemical properties of functional groups (see below), *FTrees* has the ability to detect novel molecular scaffolds. The optimum similarity of two descriptors is defined by an alignment, so an SAR may be readily detected.

FTreesXL assists the user in calculating the Feature Tree descriptor and in comparing two or more molecules in an easy and straightforward way. Also, a unique feature only found in *FTreesXL* is the clustering of molecules based on Feature Tree similarity.

2.0.2 *FTrees* and molecular similarity

Molecular similarity is one of the major concepts in the design of new drugs. This is especially the case if no three-dimensional structure of the target protein is available. It is then necessary to search for new lead structures via screening or to derive them from known inhibitors or natural substrates. Molecular descriptors are developed to make similarity searching and clustering possible for large databases of small compounds. The basic idea is to find a description of a molecule which on the one hand reflects similarity, while on the other hand enables the use of efficient comparison algorithms.

Many frequently used descriptors are based on a linear description such as bit strings or vectors. However, the Feature Tree descriptor represents the molecule by an unrooted tree in which the nodes describe the major building blocks of the molecule. The similarity of two Feature Trees is based on an alignment between the two trees. The Feature Tree approach has several advantages, the most important being a fuzzy yet topology-preserving description of the molecule, no pre-defining of fragments on behalf of the user (as in structural keys), handling of partial similarity, and the fact that the alignment of two feature trees can be translated into a comprehensible mapping of the two underlying molecules. For more details of the algorithms and results achieved see the *FTrees* User Guide.

Copyright notes

3.0.3 Legal notice and terms of use

This **material (the software and documentation package)** contains proprietary information of BioSolveIT GmbH and its subcontractors. It is protected by copyright.

The documentation is provided together with the Software of BioSolveIT under a license agreement and may be used only in accordance with the terms and conditions of this agreement. The document serves solely for the purpose of using the Software. No part of the document may be transferred to any third party or reproduced as a whole or in parts without written permission from BioSolveIT GmbH, An der Ziegelei 75, 53757 Sankt Augustin, Germany, www.biosolveit.de Phone ++49-2241-2525-0, support@biosolveit.de © 2007 BioSolveIT GmbH.

FTreesXL consists of the actual *FTreesXL* core software, a Python interface to the *COLibri* software (*PyCOLibri*), a special version of the *FTrees* software, identified by the name *FTrees-HTS*, an adaptation of the *sdg* 2D molecule drawing library and the graphical frontend. All these parts are developed by BioSolveIT GmbH. All rights reserved. © 2001-2007 BioSolveIT GmbH

THIS IS AN UNPUBLISHED WORK CONTAINING CONFIDENTIAL AND PROPRIETARY INFORMATION. DISCLOSURE, USE OR REPRODUCTION WITHOUT WRITTEN AUTHORIZATION OF THE COPYRIGHT HOLDER IS PROHIBITED.

Neither BioSolveIT GmbH nor any person acting on its behalf makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the material contained in this document or in the corresponding online documentation, or in the programs or data described herein.

BioSolveIT GmbH assumes no responsibility or liability with respect to the use of this document, any material contained herein, or any programs or data described herein, or for any damage resulting from the use of any of the above.

3.0.4 Additional copyright notes

The following software/data is used within *FTreesXL*:

- **PyCoLibri**: © 2005-2006 BioSolveIT GmbH. All rights reserved.
- **FTrees**: © 2001-2006 BioSolveIT GmbH. All rights reserved.
- **sdg**: © 2001-2006 BioSolveIT GmbH. All rights reserved.
- **Python library**: ©1991-1995 Stichting Mathematisch Centrum, Amsterdam, The Netherlands. All rights reserved¹.

¹Python copyright notice: Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appears in all copies and that both that copyright notice and this permission notice appear in supporting documentation, and that the names of Stichting Mathematisch Centrum or CWI or Corporation for National Research Initiatives or CNRI not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission.

While CWI is the initial source for this software, a modified version is made available by the Corporation for National Research Initiatives (CNRI) at the Internet address <http://ftp.python.org>. STICHTING MATHEMATISCH CENTRUM AND CNRI DISCLAIM ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL STICHTING MATHEMATISCH CENTRUM OR CNRI BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

- Qt™ © 2005 Trolltech. All rights reserved ².
- PyQt™ © 2005 Riverbank Computing Limited. All rights reserved ³.
- SIP © 2005 Riverbank Computing Limited. All rights reserved ⁴.

3.0.5 Trademarks

The registered trademark or trademarks of BioSolveIT GmbH include but are not limited to: BioSolveIT™, FlexX, FlexS, FlexV, FlexE, FlexX^C, FlexX-Pharm, CoLibri, FTrees, HTSView, FTreesXL, PyFlexX, PyFlexS, PyFTrees, PyCoLibri, DDB, and FlexDB.

- SMILES™ and SMARTS™ are registered trademarks of Daylight Chemical Information Systems.

All other names of products and companies are used for reference only. The omission of a “TM” label does not imply that the terms can be used freely. They may be trademarks of the respective owner.

²Qt: The development is based on a Qt Commercial License.

³PyQt: The development is based on a commercial license.

⁴SIP: SIP is licensed under the same terms as Python itself. SIP places no restrictions on the license you may apply to the bindings you create.