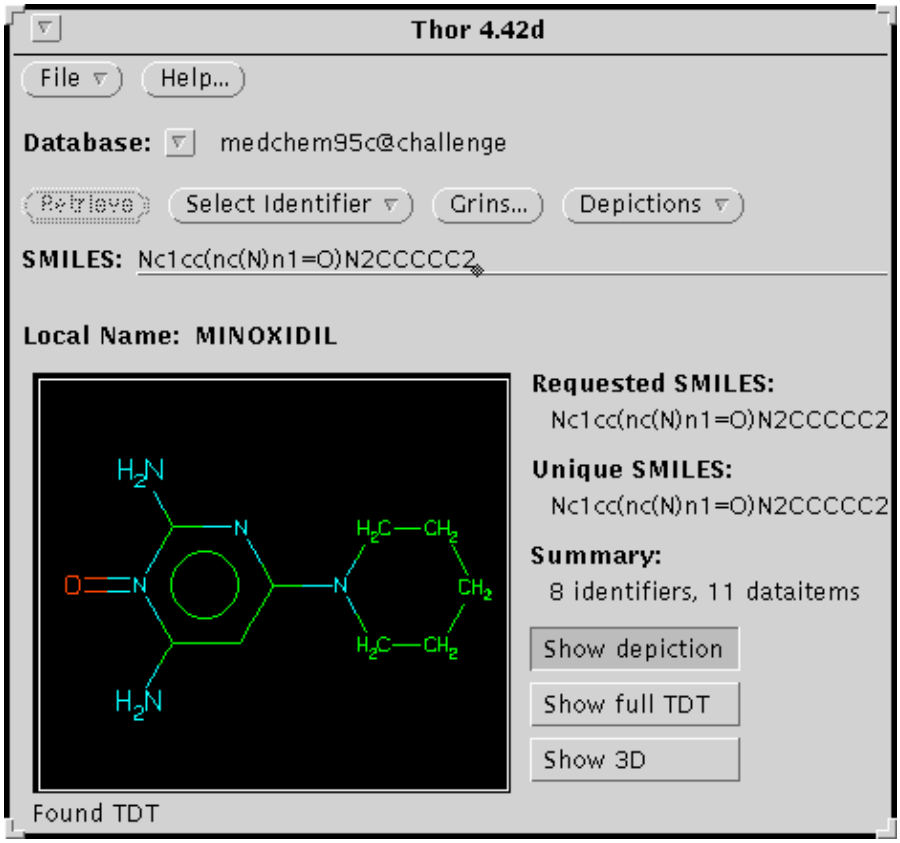
Daylight Chemical Information Systems, Inc.
120 Vantis - Suite 550 - Aliso Viejo, CA 92656
tel +1 949-831-9990 - fax +1 949-831-9902 - www.daylight.com

XVTHOR Manual

Table of Contents

XVTHOR Manual

Daylight Version 4.9
Release Date 08/01/11



1. Introduction to XVTHOR

THOR (THesaurus Oriented Retrieval) is a database designed to store and retrieve chemical information in an efficient, rational, and convenient manner. THOR is designed specifically for chemical information processing; the primary key used in the database is the molecular structure as defined by its SMILES.

Extremely fast data retrieval time, independent of database size, is achieved through hash-table algorithms. Look-up time will be constant and fast for both small databases of a few structures or large databases of exceeding 10 million structures. Space efficiency is attained by the minimal storage requirements of the hash algorithm and SMILES compactness.

As of version 4.0, THOR is a client-server system for a network environment. Servers, clients, and databases can be distributed throughout a computer network. This approach offers many advantages in flexibility and performance; it is well suited to modern network environments. The system is transparent to a typical user;

simultaneous access to one or more databases is provided in a THOR-client window.

This XVTHOR User Guide is intended to provide sufficient information to use the THOR client program XVTHOR. See the Daylight System Administration Manual for information on managing a THOR/Merlin database system. For programming information, refer to the THOR Toolkit Programmer's Guide.

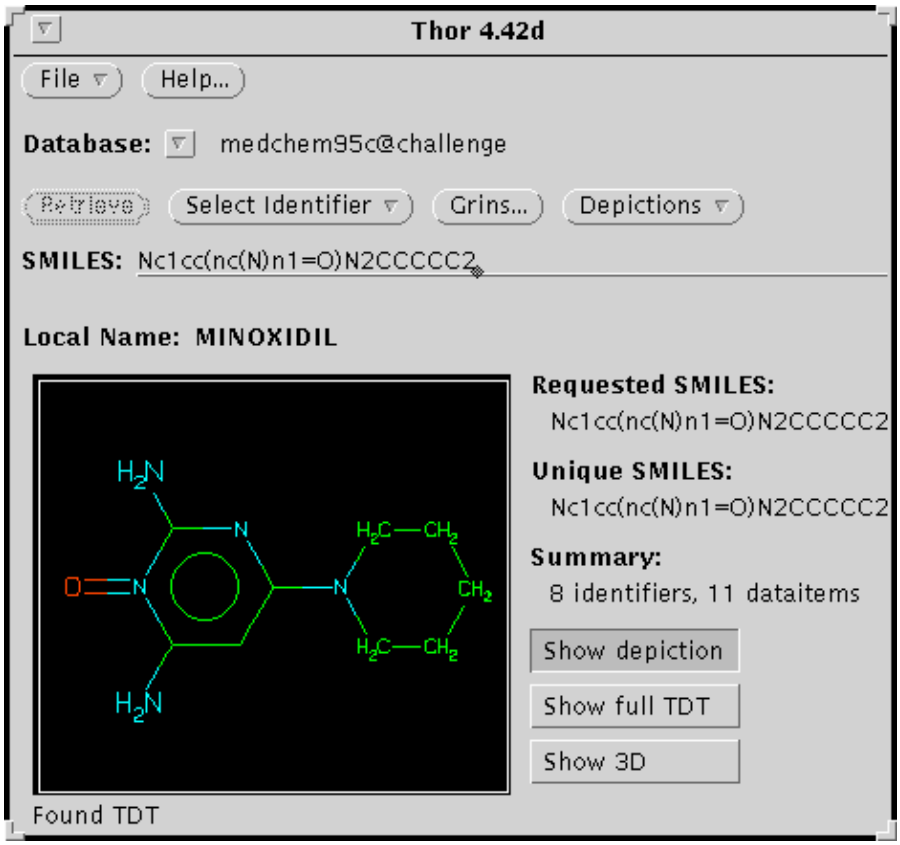
2. Getting started

Prerequisites for running XVTHOR:

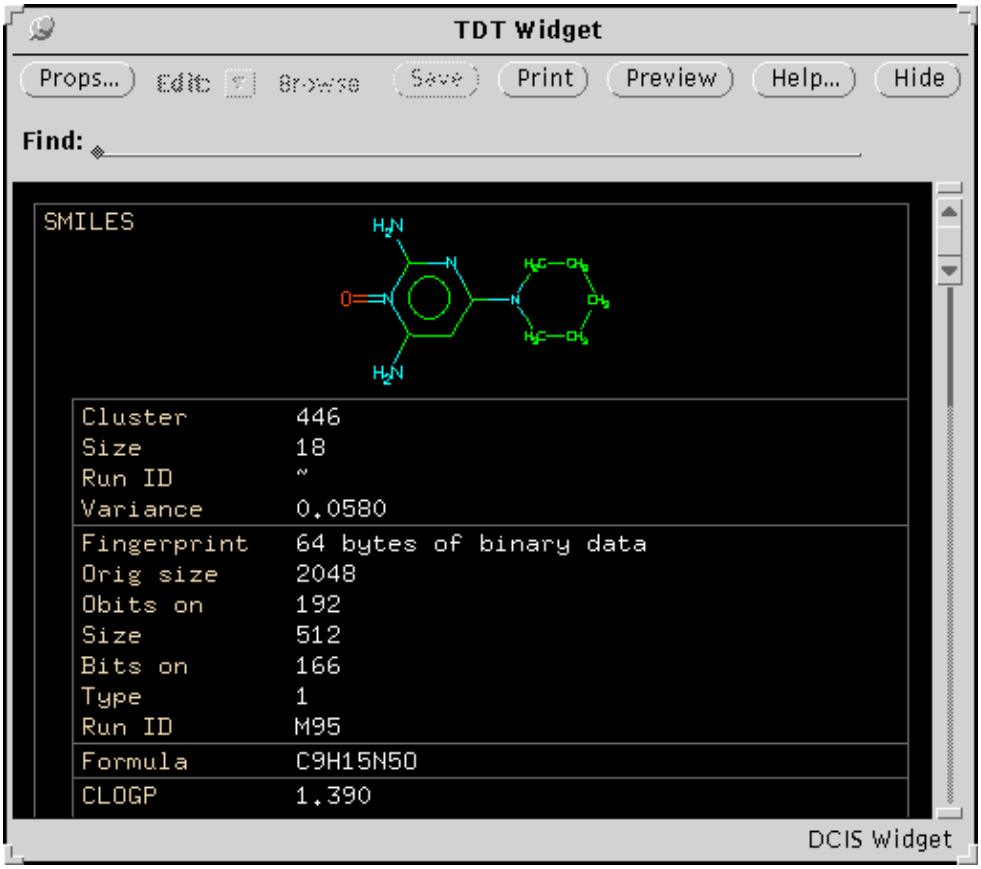
- The THOR program has been installed locally.
- A database has been installed and is accessible to the server.
- The THOR-server has been started.
- The client is permitted access by the THOR-server.
- Local environment variables have been defined (normally **DY_ROOT** and **DY_LICENSEDATA**.
- The Daylight Software License is valid for "thor".
- To start the THOR program enter: "xvthor" (for SGIs use "xvthor4d").

3. Basic operation of XVTHOR

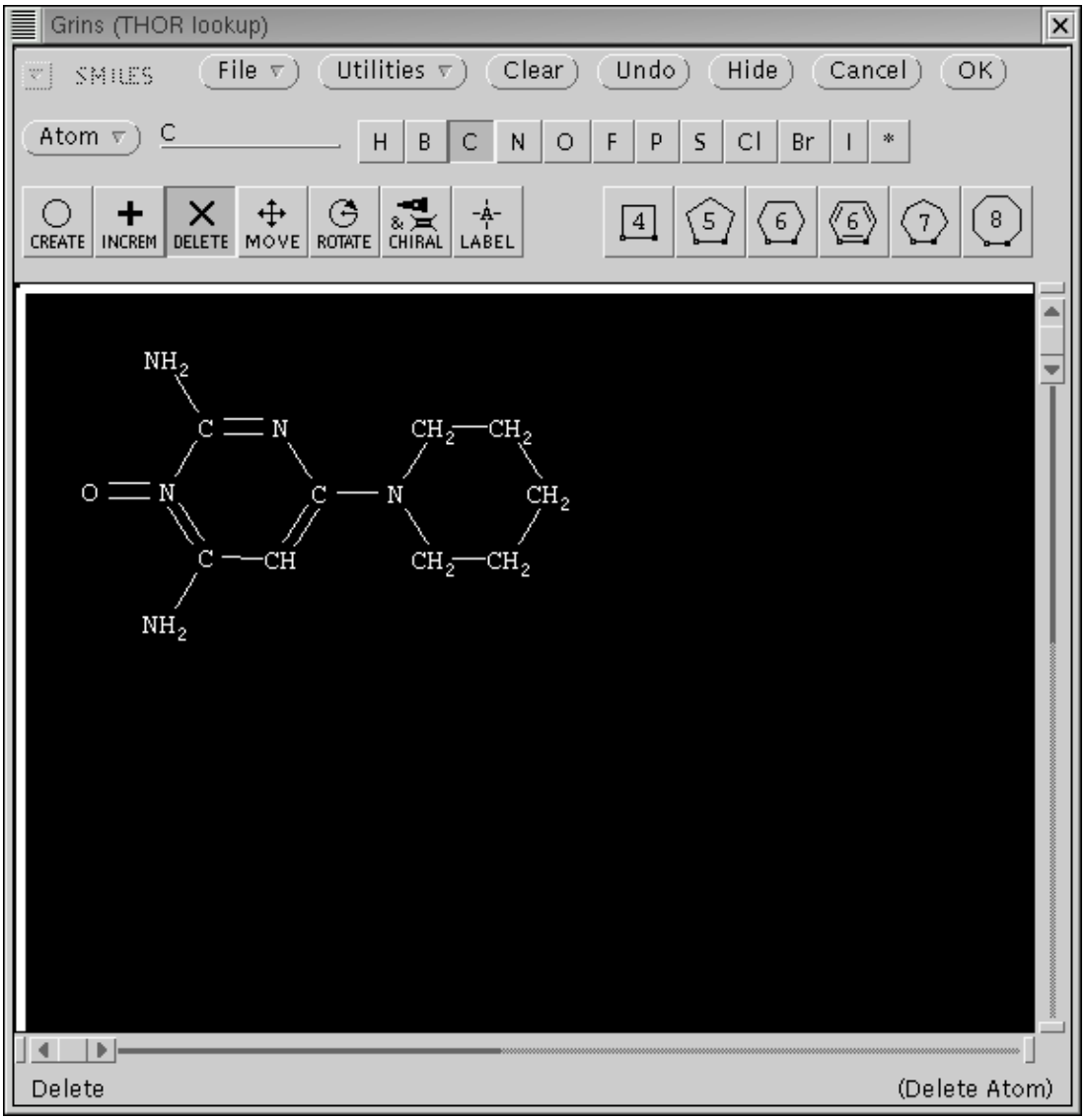
In the X-windows environment, THOR appears to the user as a set of windows. These windows are as shown below:



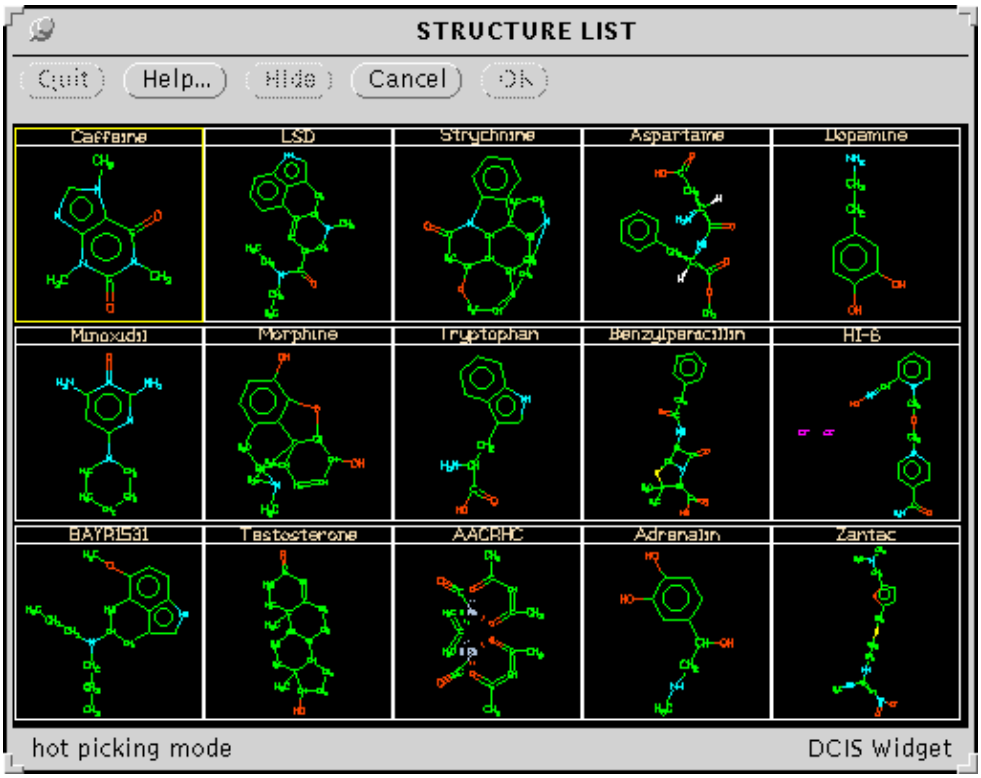
The main THOR window: Use this window to enter a SMILES, NAME, or other identifier directly. Status information and a depiction are displayed.



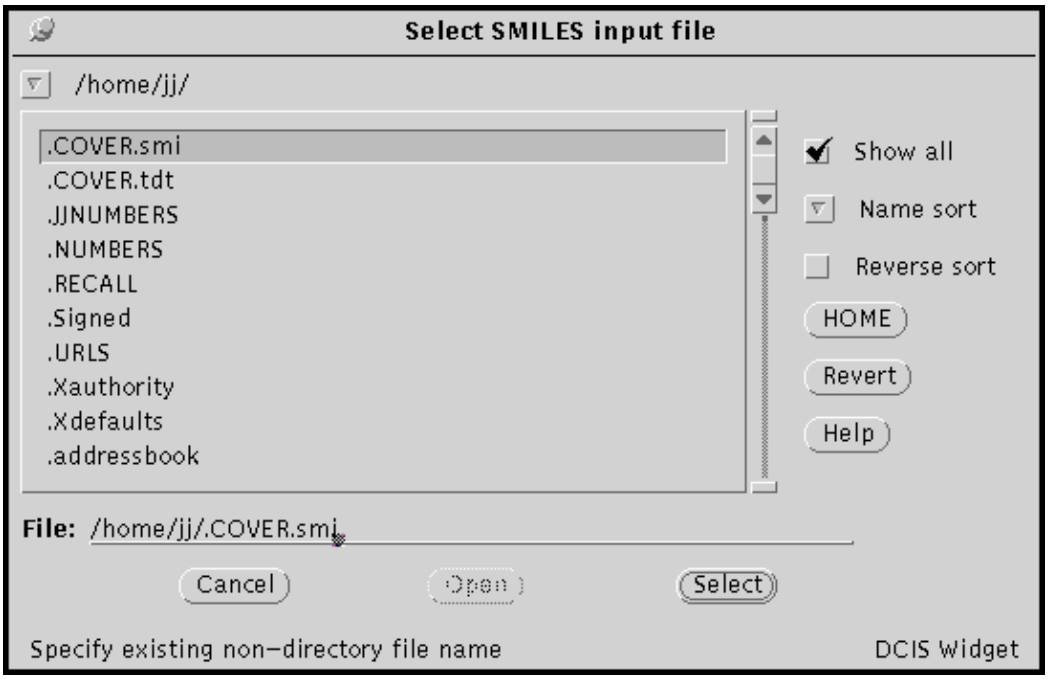
THOR Data Tree (TDT) Display Window: This window displays the contents of the database for the designated structure. Also known as the "TDT Widget."



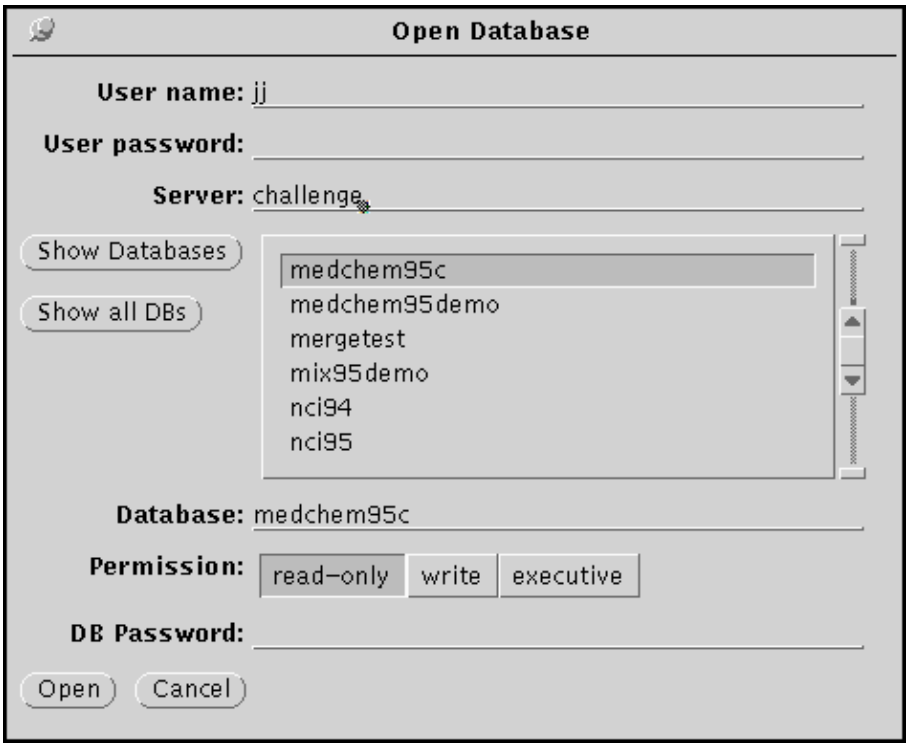
GRINS Window: GRINS allows graphical input of a structure. Build a molecule from atoms, bonds, templates, and parent structures, then select it for THOR lookup. Also known as the "GRINS Widget."



Structure List: Display the contents of a SMILES file for graphic selection of a structure. Blow up one structure with the middle mouse button for closer examination. Also known as the "Depict Widget."



SMILES File: Specify a SMILES file for a stored list of structures.



Open database: Opens a database by connecting with the THOR server.

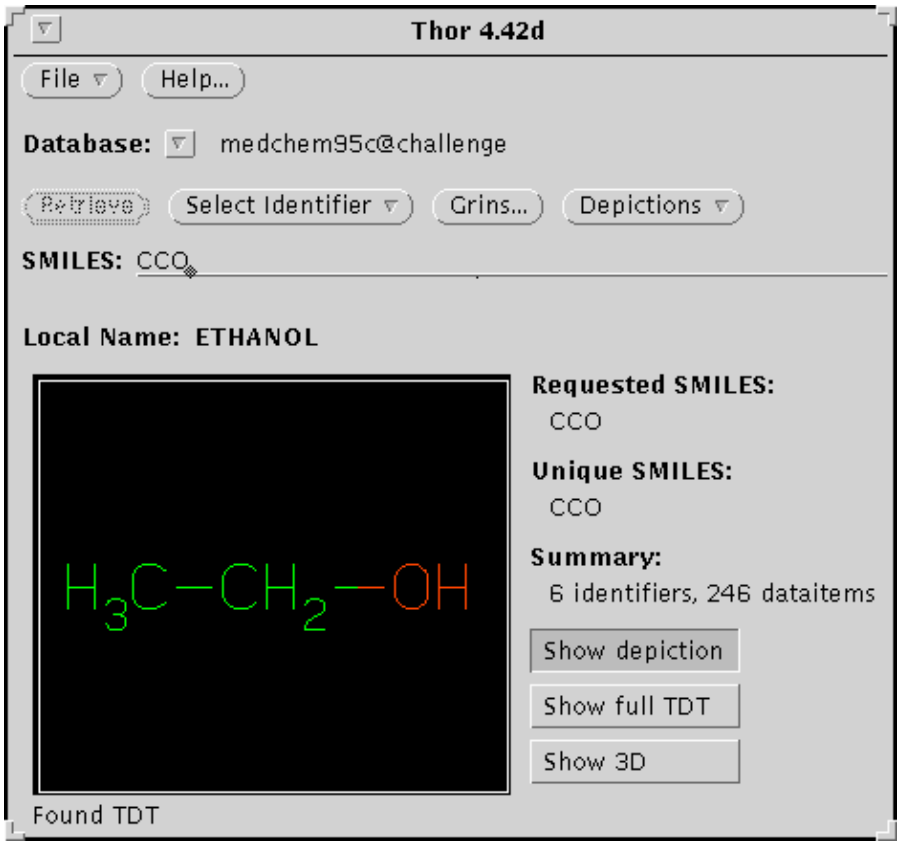
3.1 Starting XVTHOR

To start the THOR-client program, just type xvthor (for XView THOR). A window should appear on the desktop which is the main THOR window and allows access to all other THOR menus, windows, and functions.

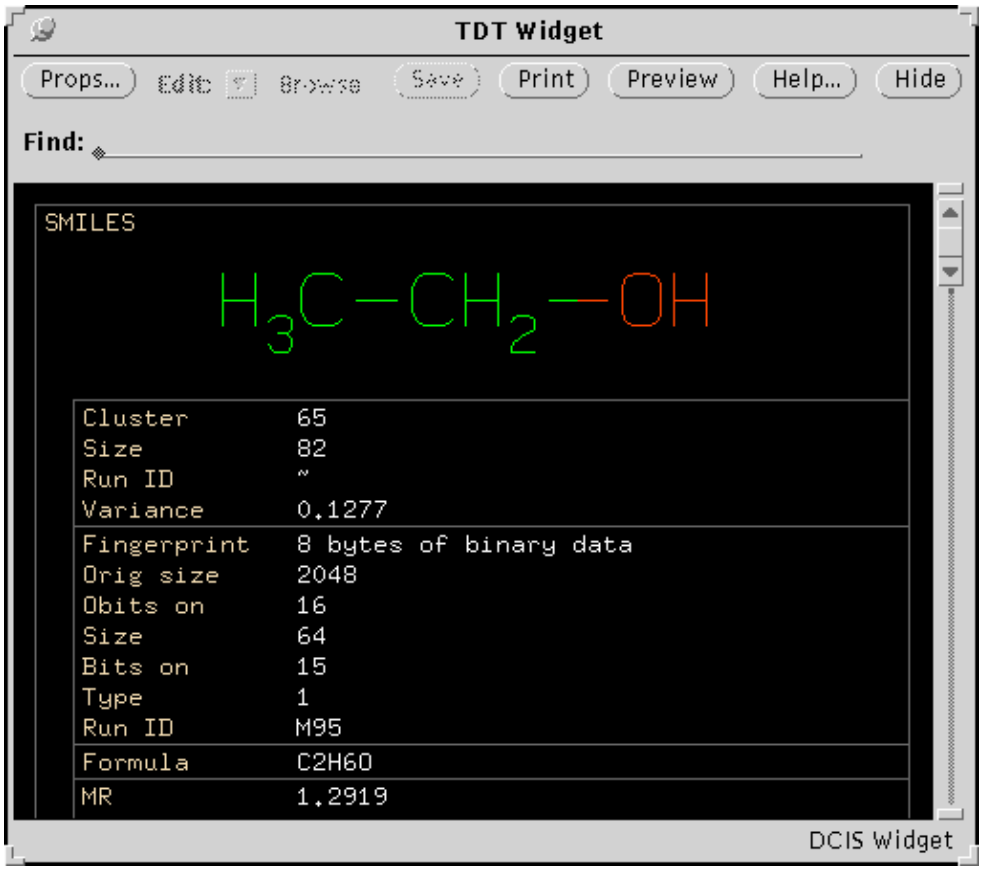
3.2 Opening a Database

To open a THOR database, position the mouse cursor on the Database button and press the menu mouse-button. Drag the mouse to Open Database and release.

3.3 A Simple Query



To look up data for ethanol, position the cursor on the input line and press the select mouse-button. Then type the SMILES "CCO" on the input line and press <return>.



The THOR window showing SMILES "CCO" requested." If ethanol is present in the currently opened database, THOR will provide a TDT window to display the data.

4. Lookup Operations

An identifier is a datatype which may be used to look up data (e.g. name, registration number, SMILES). Data associated with a chemical structure is stored with that structure in a SMILES-rooted data tree. The ordering of data in this tree specifies data relationships. Data items which are not identifiers belong to preceding identifiers. Identifiers belong to the SMILES that is the root of the THOR data tree(TDT). This ordering is shown in the TDT display with lines and indentation. The THOR client allows structures to be looked-up by any associated identifier.

4.1 Lookup by SMILES

SMILES may be entered lexically or graphically. To lexically enter a SMILES: select SMILES as the input datatype, type the SMILES on the input line, and press <return>. To specify a structure graphically: press the GRINS button with the select mouse-button, enter a structure, and press the SELECT button. If the SMILES is present in the current database, the THOR window will indicate that the TDT has been found, the structure will be depicted, and the number of data items will be displayed. If the SMILES is not present, this will be indicated. The GRINS widget is described in detail in the Daylight Widgets User's Guide.

4.2 Lookup by Other Identifiers

To look up data using a non-SMILES identifier, select the input datatype (e.g. NAME), type the datatype value (e.g. CAFFEINE) on the input line, and press <return>. If no TDT is found, this will be indicated. If one TDT is found with the specified identifier, this data will be retrieved. If more than one TDT is found with the specified datatype value (i.e. ambiguous identifier entered), depictions of each structure will be displayed. One structure can be selected from those displayed.

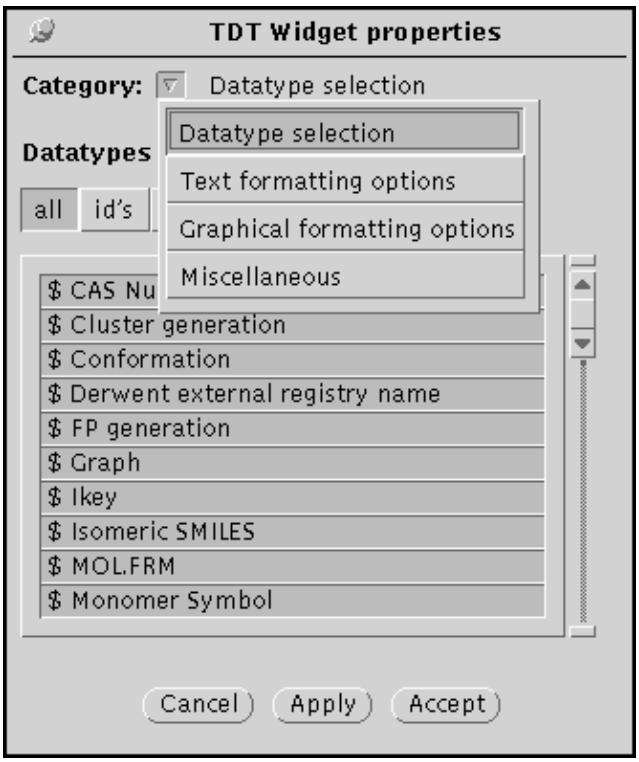
4.3 Identifier Standardization

Most identifiers are standardized to improve lookup effectiveness. NAMEs are shifted to uppercase and have blanks and non-alphanumerics removed. When the name methanamine, N-hydroxy is entered, it is interpreted as METHANAMINENHYDROXY. SMILES are uniquified; for any given structure all possible input SMILES map to the same unique SMILES.

4.4 Non-SMILES Root Identifiers

THOR can store data for substances without known structures. Since no SMILES is present, the TDT is "Non-SMILES rooted". Non-SMILES rooted TDTs are retrieved by their root ID, in the usual manner. At the present time (4.3), crossreference identifiers are not allowed for non-SMILES rooted TDTs.

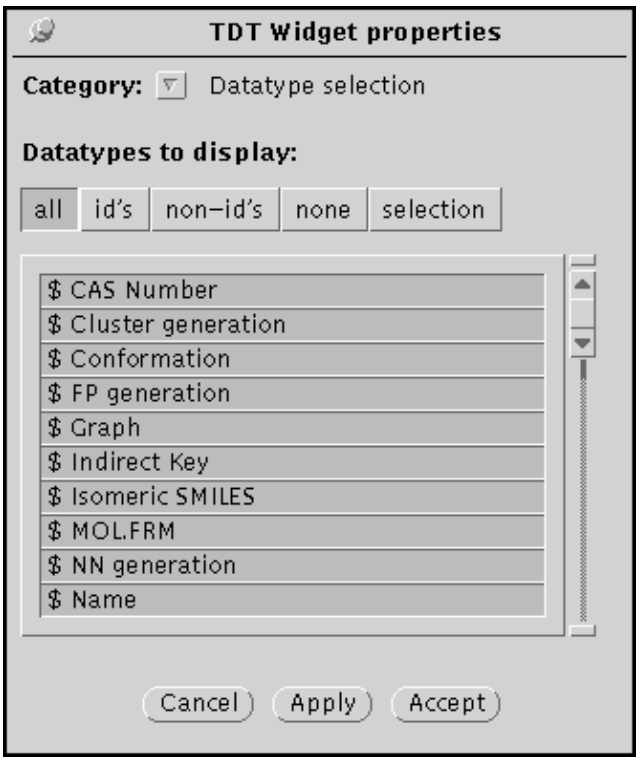
5. Data Display



THOR displays the TDT for a structure by using the TDT Widget. The lines and indentation represent the three-level tree structure of the data: the root identifier (usually SMILES), identifiers associated with the

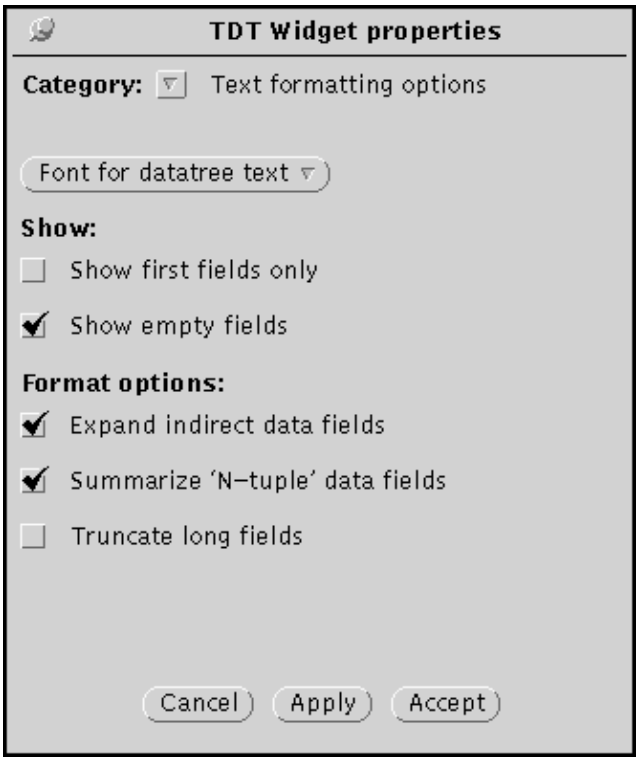
root-ID, and data items which are associated with the nearest preceding identifier. Press the Props... button on the TDT window to invoke the TDT Widget properties panel:

The TDT display may be modified in several ways by the properties window, which can appear in one of four modes: Datatype selection, Text formatting, Graphical formatting, and Miscellaneous.

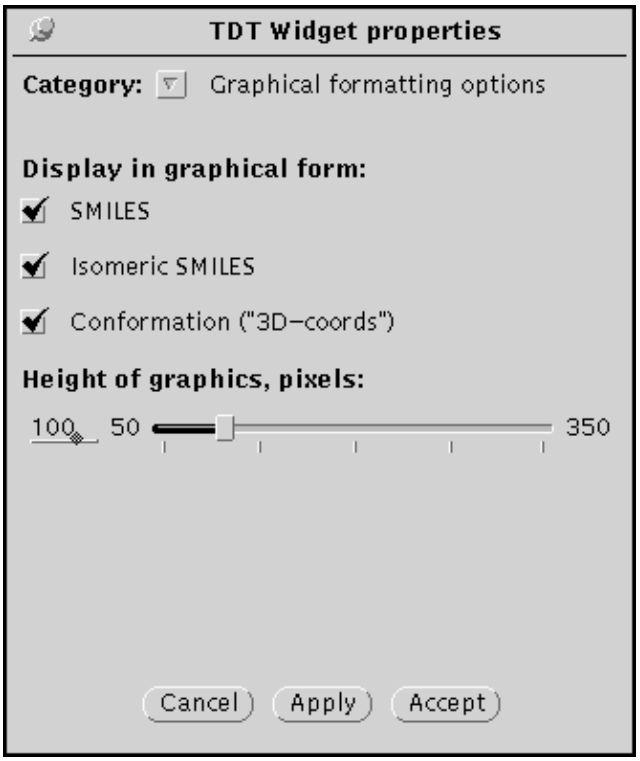


The datatype selection panel specified which datatypes are displayed by the TDT Widget. Identifiers are indicated by a dollar sign ($) in the scrolling region.

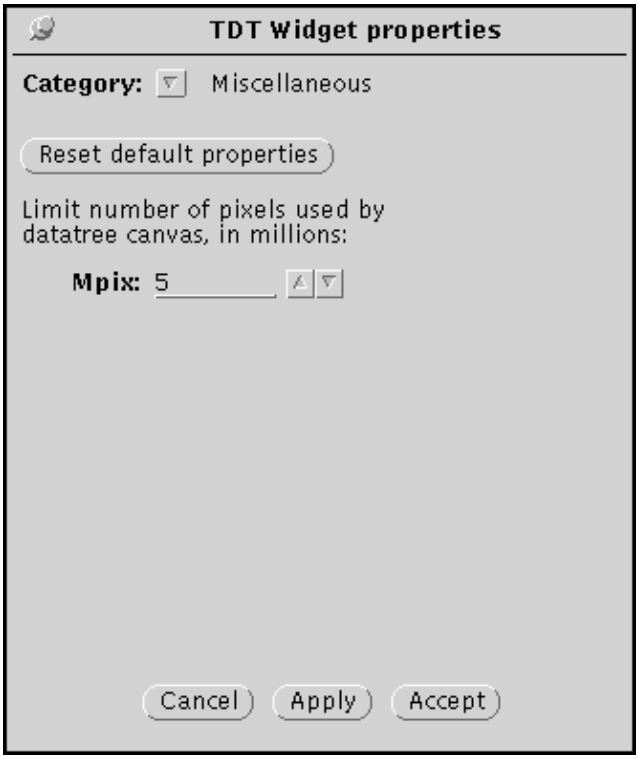
Note that the 'Apply' button implements the specified choices; 'Accept' also hides the window



The text formatting panel specifies the appearance of dataitems and their fields in the TDT Widget.



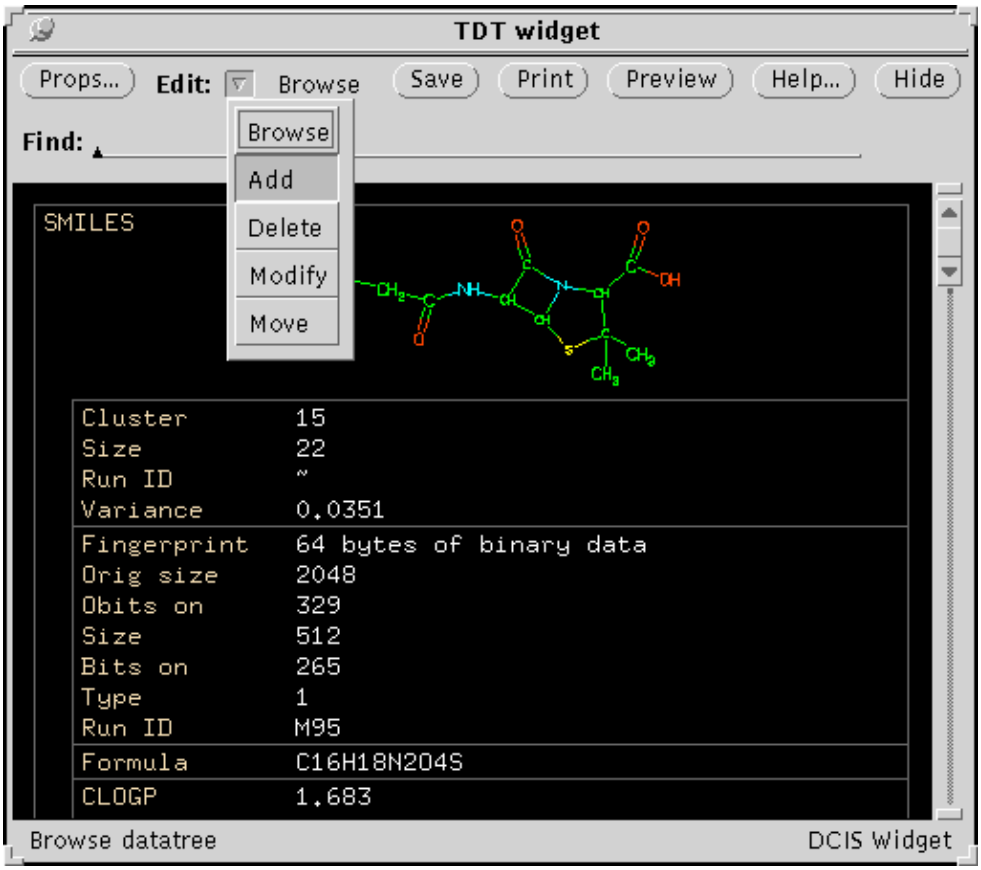
The graphical formatting panel specifies which data are to be displayed graphically. The slider specifies the size of the graphics.



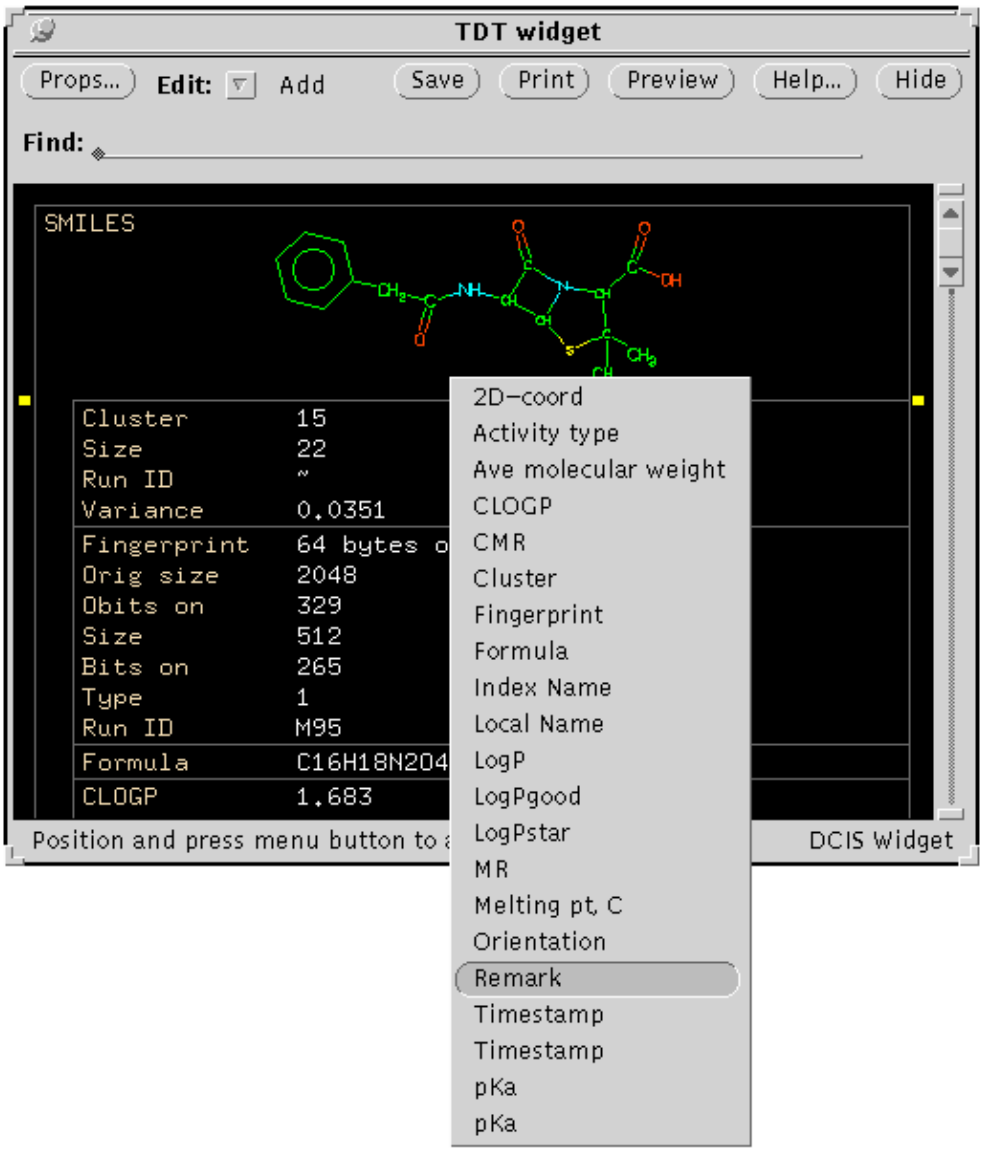
The miscellaneous panel specifies the maximum size of the TDT Widget canvas.

If a TDT does not fit in the allocated canvas, lines are drawn to show the location of missing data.

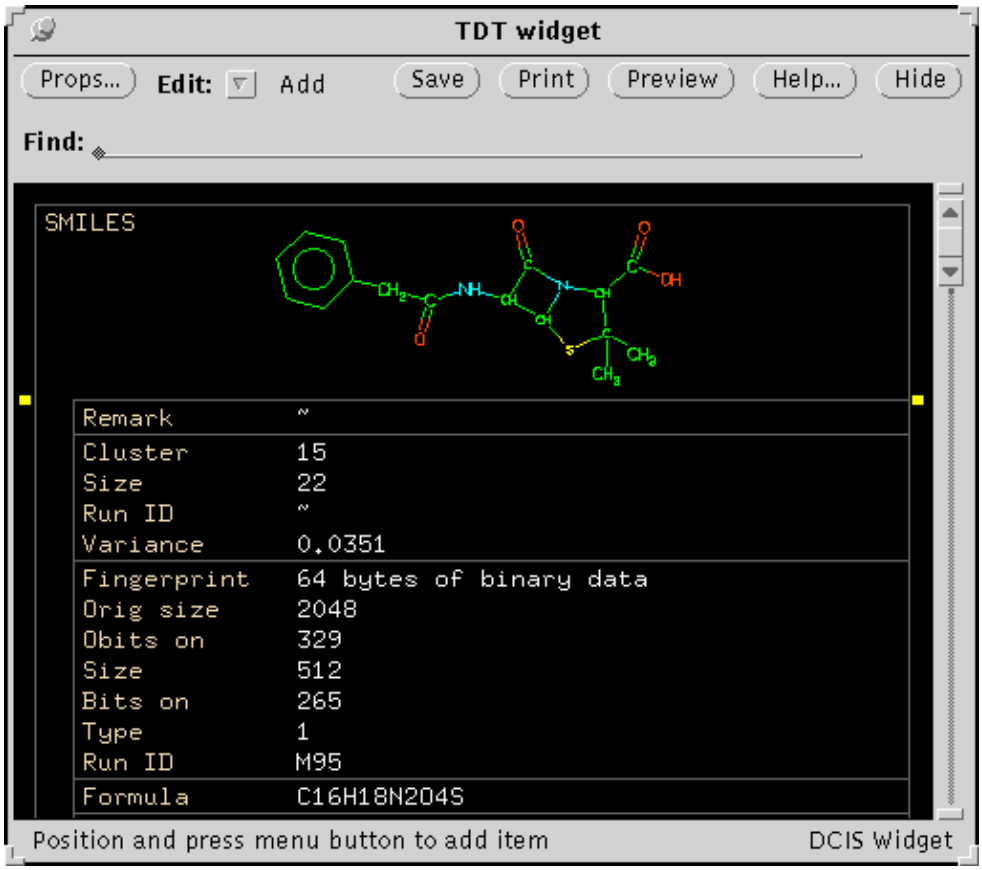
6. Write Operations



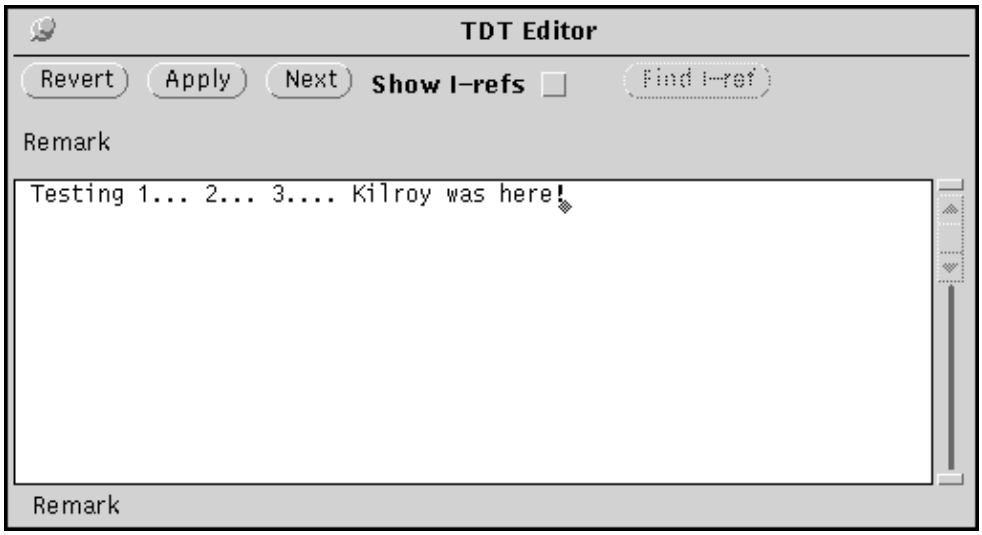
The THOR Datatree window initially is in a read-only "Browse" mode. The database must be opened in "write" permission to change the contents of the TDT. To add new data, position the mouse cursor on the "Edit" button, press the menu mouse-button, and drag the mouse to "Add":



Lines where dataitems may be added will be highlighted. Use the mouse to select a location for the new dataitem. Press and hold the mouse menu-button. A list of possible datatypes will appear; drag the mouse to the datatype to be added and release:



A new dataitem will be added with empty datafields. The contents of a dataitem may be added or changed in "Modify" mode. Position the mouse cursor on the "Edit" button, press and hold the menu mouse-button, drag the mouse to "Modify", and release. Datafields that may be modified will be highlighted:



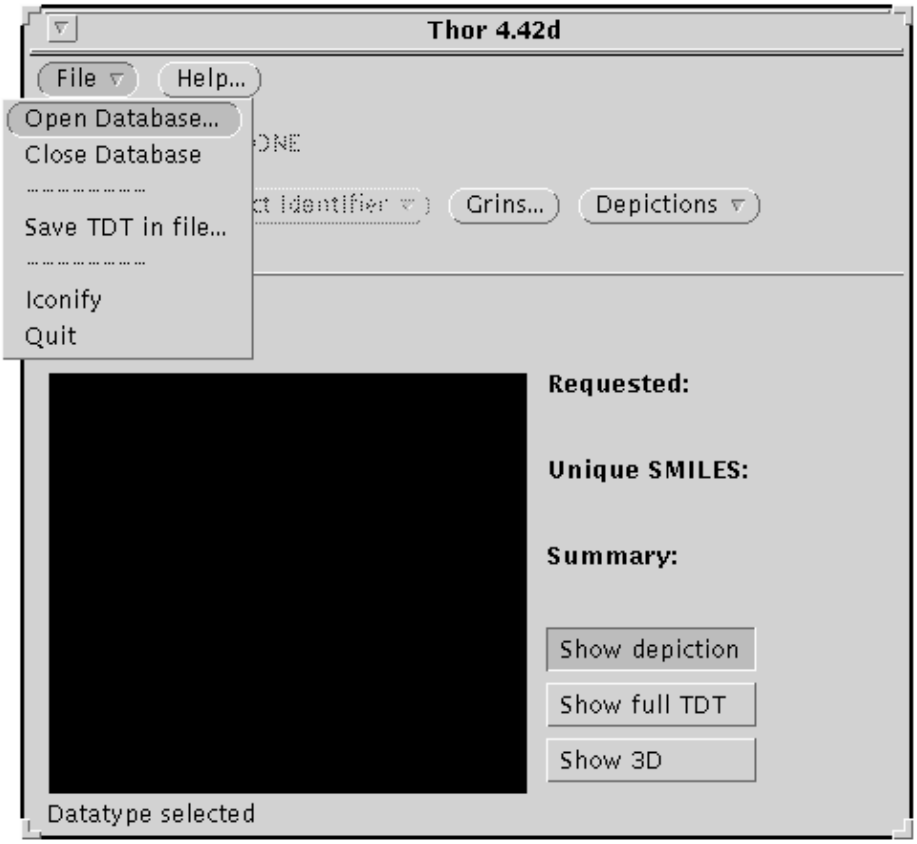
Use the mouse to select the dataitem to be modified. Pressing the select mouse-button will cause a "Text Editor" window to appear. Enter new data or alter current data, press the "Apply" button to save changes, or the "Revert" button to ignore changes.

Dataitems may be deleted by selecting "Delete" in the "Edit" pull-down menu. Dataitems may be moved by selecting "Move" in the "Edit" pull-down menu.

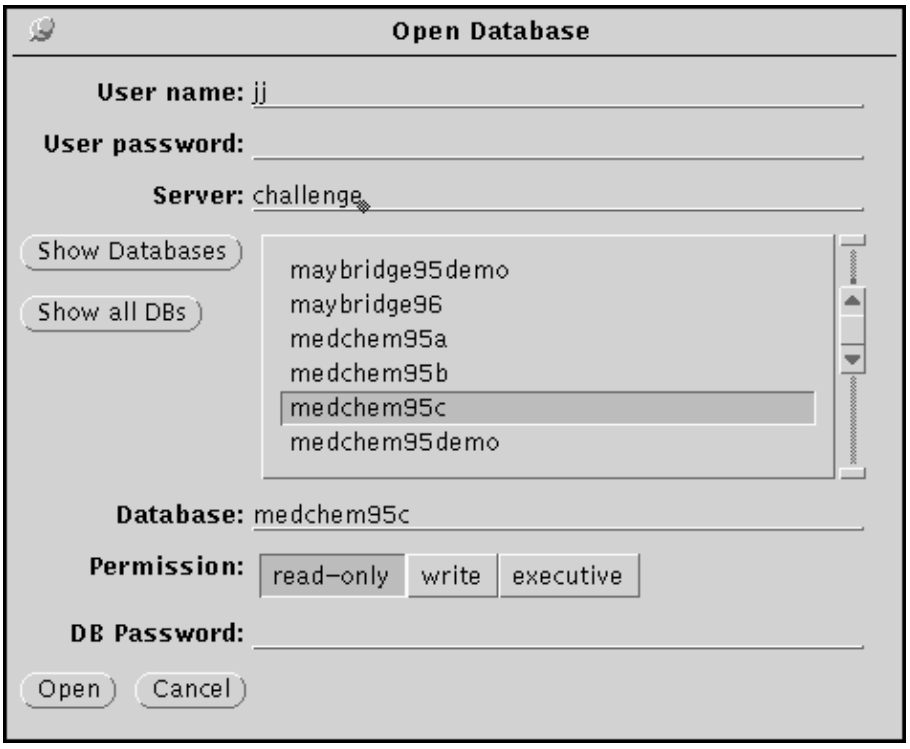
7. On-line Help

Press the "Help" button on the THOR window to invoke the Daylight Help Widget.

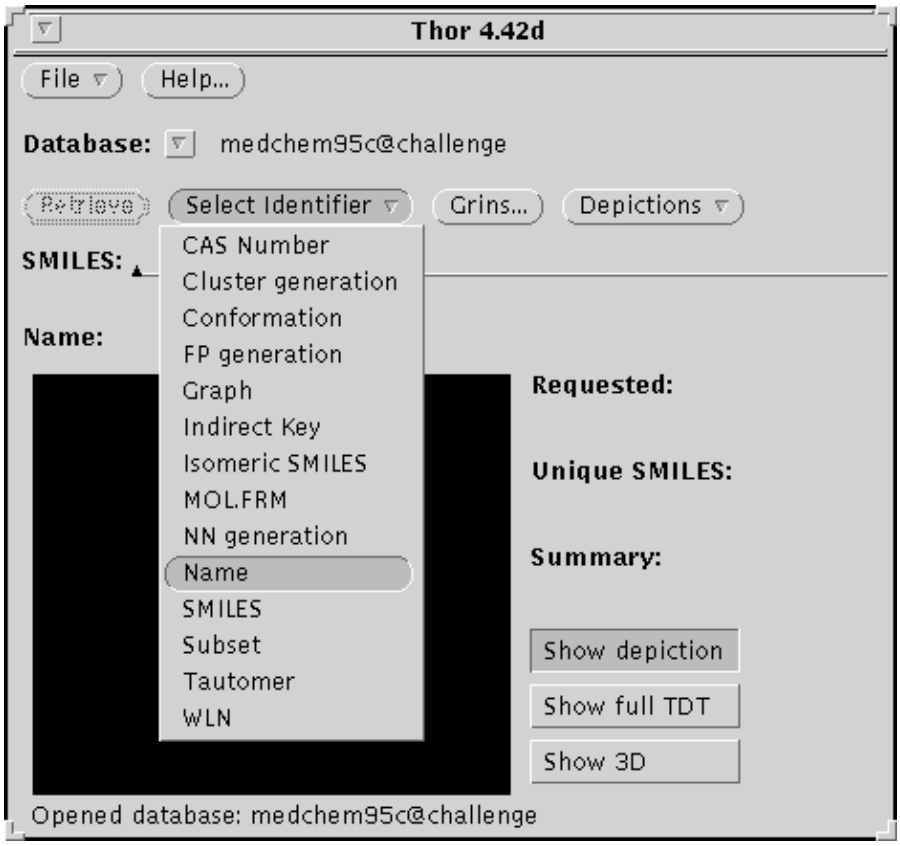
8. Sample Session



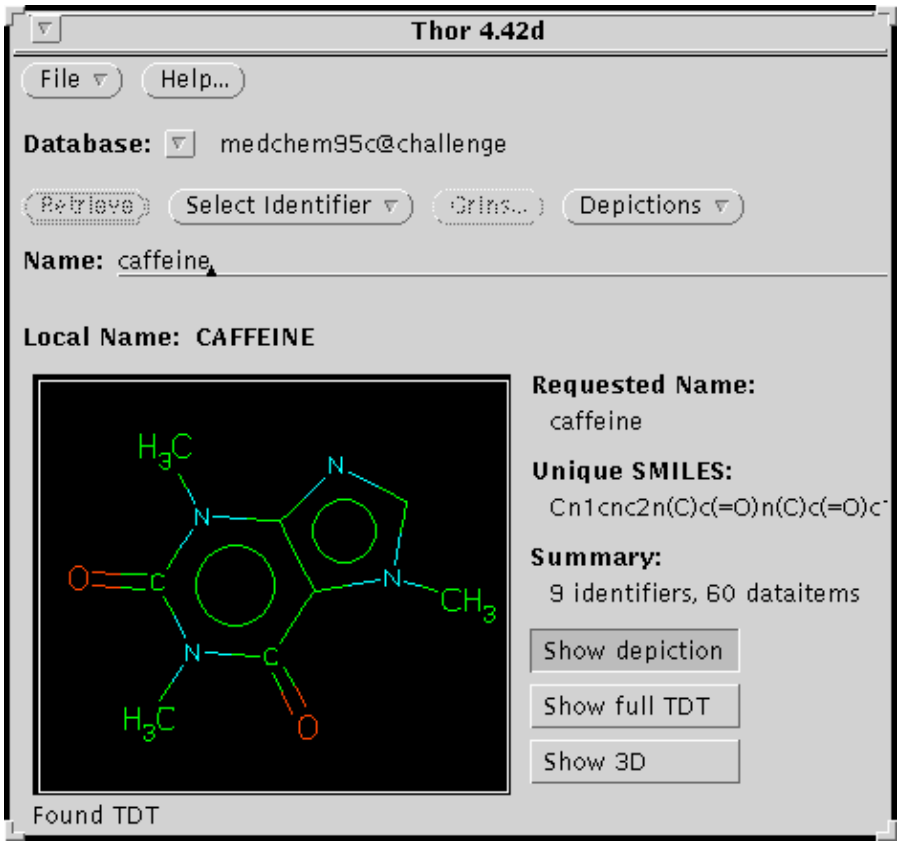
A THOR server must be running before the THOR client can access a database. See the System Administration Manual for more information on starting a THOR server. After starting THOR, the next step is to open a database. Use the menu mouse-button to choose "Open Database" from the Database menu.



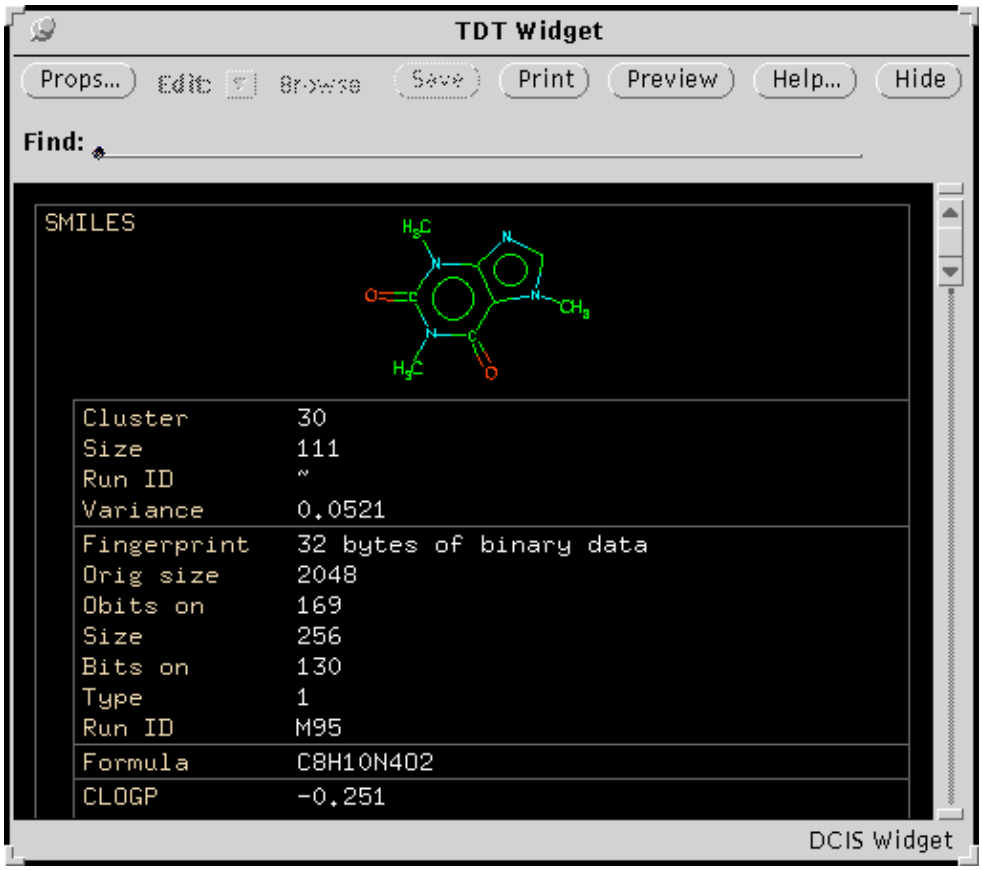
Type the database name and password (if any) into the appropriate fields. Click the Open button to open the database. Note that if the THOR server is on another machine, the machine name precedes the database name, separated by a colon.



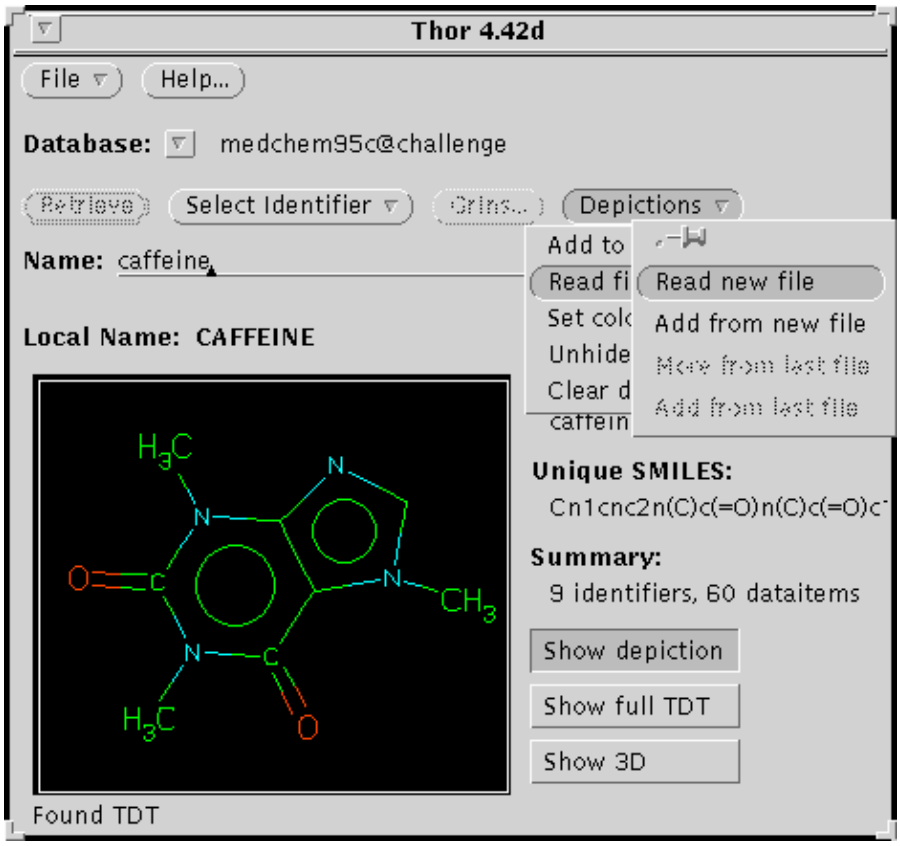
Once the database is open, look up a structure by name. Change the datatype to NAME by selecting from the Select Datatype menu:



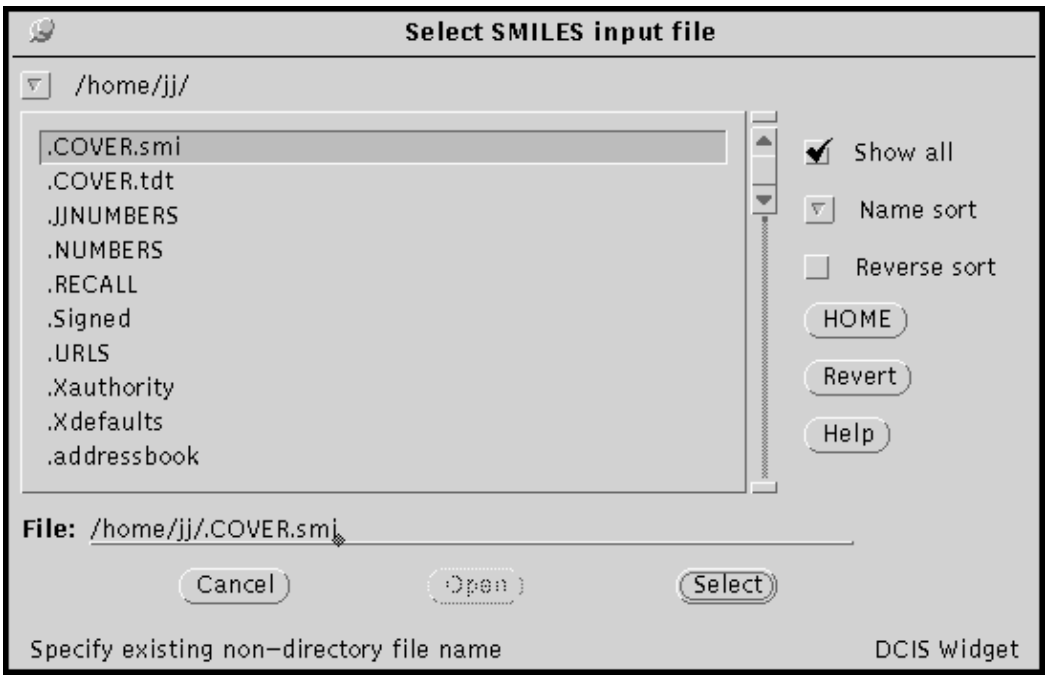
Type the structure name on the input line and press <return>. If you are using the Daylight demo database or Pomona, try "caffeine."



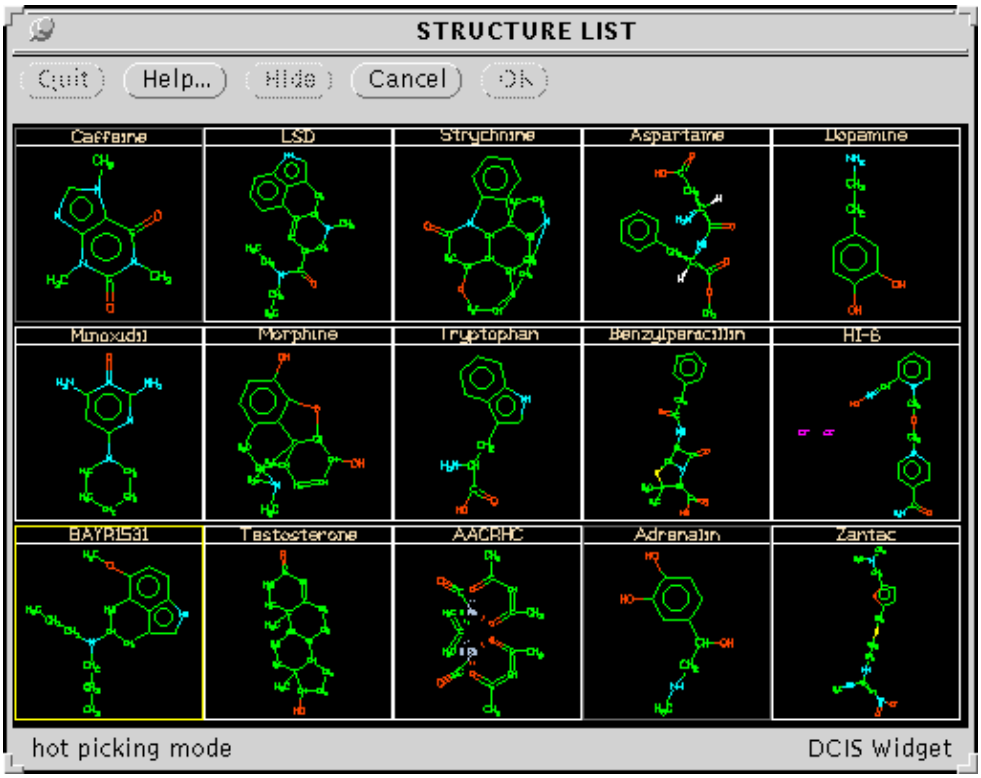
If the structure is present in the database, a TDT window will appear containing all data.



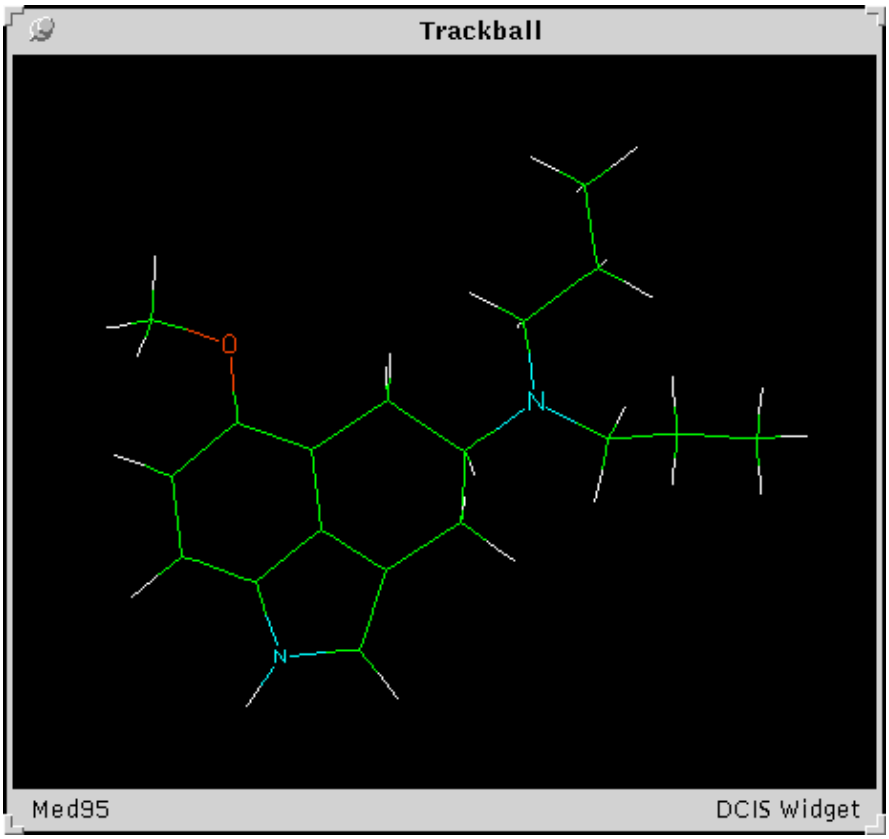
A structure may be graphically selected from a SMILES file. Use the mouse menu button to choose "Read new file" from the Depictions menu.



This will bring up a file-selection panel. Find a SMILES file and press "Select."



A window of depictions will appear. With this window, using the Select mouse button results in a THOR lookup, and using the Adjust mouse button results in a blowup window of the selected structure:



The 3D display widget, or "Trackball Widget."

9. Using THOR with other Daylight software

The Daylight database system is comprised of two distinct parts: THOR and MERLIN. THOR is designed for the efficient storage and retrieval of chemical data (i.e. data lookup). MERLIN is designed for substructure-, similarity-, and string-searching. The amount of data that can be searched is limited by the memory of the machine, since MERLIN utilizes in-memory searching for speed. Designing a THOR/MERLIN system involves deciding what data to select for MERLIN based on searching needs and available memory resources. Typically, MERLIN and THOR will be used with the same set of structures, and the MERLIN data will be a subset of the THOR database. Using the window interface, it may be convenient to use MERLIN and THOR simultaneously, searching for structures of interest with MERLIN and then looking up their complete data trees with THOR.