



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

Daylight XVMerlin Manual

Table of Contents

<u>XVMerlin Manual</u>	1
<u>1. Introduction to XVMerlin</u>	1
<u>2. Basic Operation of XVMerlin</u>	2
<u>3. Using the XVMerlin Window Menus</u>	4
<u>3.1 The Hitlist Menu</u>	4
<u>3.2 The Display Menu</u>	5
<u>3.3 The Keypad</u>	6
<u>3.4. The File Menu</u>	6
<u>3.5 The Data Column Menu</u>	8
<u>3.5.1 Draw Hits & the Depict Widget</u>	10
<u>3.6 The Popup Menu</u>	11
<u>4. Searching</u>	13
<u>4.1. Sample Searches</u>	16
<u>4.1.1 A String Search</u>	16
<u>4.1.2 A Superstructure Search (SMILES)</u>	18
<u>4.1.3 A SMARTS Search</u>	19
<u>5. Configuring Merlin</u>	21
<u>Saving and restoring XVMerlin's state</u>	22

XVMerlin Manual

Daylight Version 4.9
Release Date 08/01/11

The screenshot shows the XVMerlin software interface. At the top, it says "Merlin 4.42d - DCIS (c) 1994". Below that are menu buttons: File, Hitlist, Display, Help..., and Messages... The database is set to "spresi95preps@challenge:merlin:jj". The current position is 2, with 406 hits and a total of 983839. There are buttons for "SMILES", "article", and "art/author". The main display area is a table with five rows, each containing a chemical structure on the left, a text description in the middle, and author names on the right. The first row shows a complex organic molecule and describes the formation of an angular 6-chlorobenzene-indole derivative. The second row shows a Fischer carbene complex and describes its interaction with acetylenes. The third row shows a Fischer indolization product and describes the reaction of ethyl phenylpyruvate. The fourth row shows a cyclohexadienone and describes its annulation via Fischer carbene complexes. The fifth row shows a thermolysis product and describes the Fischer-Tropsch step. At the bottom, a message says "Removed item from hitlist."

	FORMATION OF AN ANGULAR 6-CHLOROBENZENE-INDOLE BY THE FISCHER INDOLIZATION OF A 1-METHOXY-2-NAPHTHYLHYDRAZONE: CORRECTION OF THE STRUCTURE OF THE SO-CALLED LINEAR	ISHII HISASHI, MURAKAMI YASUOKI, WATANABE TOSHIKO, IWAZAKI AYAKO, SUZUKI HIDEHARU, MASAOKA ASAKA TOHRU, MZUMA YUKA
	INTERACTION OF FISCHER CARBENE COMPLEXES OF THE TYPE W(CO) ₂ C(OM)E)R> (R = CH(3), C(2)H(5)) WITH ACETYLENES	MACOMBER D. W.
	FISCHER INDOLIZATION AND ITS RELATED COMPOUNDS. XVIII. FORMATION OF FOUR UNEXPECTED ANGULAR BENZENE-INDOLES ON FISCHER INDOLIZATION OF ETHYL PHENYLPYRUVATE 2-(1,4-DIMETHYL-5-PHENYL-1,3-BUTADIENE-1-CARBOXYLIC ACID)	ISHII HISASHI, MURAKAMI YASUOKI, WATANABE TOSHIKO, SUZUKI HIDEHARU, YASUDA ZENKICHI, IKEDA NISABURO, MITSUI HIDEKI, TANI SHOHEI
	CYCLOHEXADIENONE ANNULATION VIA 'ALPHA',BETA'-UNSATURATED FISCHER CARBENE COMPLEXES	TANG PENG-CHO, WULFF W. D.
	THERMOLYSIS AND PHOTOLYSIS OF (ET)A'-C(5)MES)OS(CO) ₂ CH(2)OH: A RETRO FISCHER-TROPSCH STEP	MAY C. J., GRAHAM W. A. G.

Copyright notice

This document is copyrighted © 1991-2011 by Daylight Chemical Information Systems, Inc. Daylight explicitly grants permission to reproduce this document under the condition that it is reproduced in its entirety, including this notice. All other rights are reserved.

1. Introduction to XVMerlin

Merlin is designed for chemical database searching, and is particularly good at structural searching. The inherent compactness of SMILES in conjunction with a structural "fingerprint" allow Merlin to maximize searching speed by searching in memory. A database of 100,000 structures might typically be searched in 10MB of memory.

Merlin's capabilities complement those of THOR, Daylight's disk-resident database system. With THOR able to store very large datasets, and provide fast retrieval and read/write capability for one THOR Data Tree at a

time, it is Merlin's job to keep a subset of a THOR database in memory for fast searching. Daylight has separated the functions of data lookup (THOR) and data searching (Merlin) to optimize the performance of both. However, the Merlin client can access the Thor server to lookup the datatree for a xvmerlin row.

Merlin provides flexible display of text, 2D and 3D graphics, and allows structure input via SMILES or graphic entry (GRINS). Substructure, similarity, and string searching are available, as well as a flexible sorting menu for numerical and textual data.

This XVMerlin User Guide is aimed at the beginning user. It describes and illustrates the main capabilities of Merlin, and should enable the user to start searching a database. Refer to the Daylight THOR and Merlin Administration Guide for more information on the methodology of Merlin. Prerequisites for running Merlin:

- The Merlin program has been installed locally
- A database has been installed and is accessible to the server.
- The Merlin server has been started, and a database "pool" has been loaded.
- Local environment variables have been defined (normally **DY_ROOT** and **DY_LICENSEDATA**).
- The Daylight Software License is valid for "merlin".
- The read password for the database and server password are known (if any).
- To start the Merlin program, enter: "xvmerlin" (for SGIs, "xvmerlin4d").

2. Basic Operation of XVMerlin

The Merlin client, xvmerlin, appears to the user as a set of windows. The main Merlin window accesses all other windows and menus, and displays status information and a configurable set of data columns.

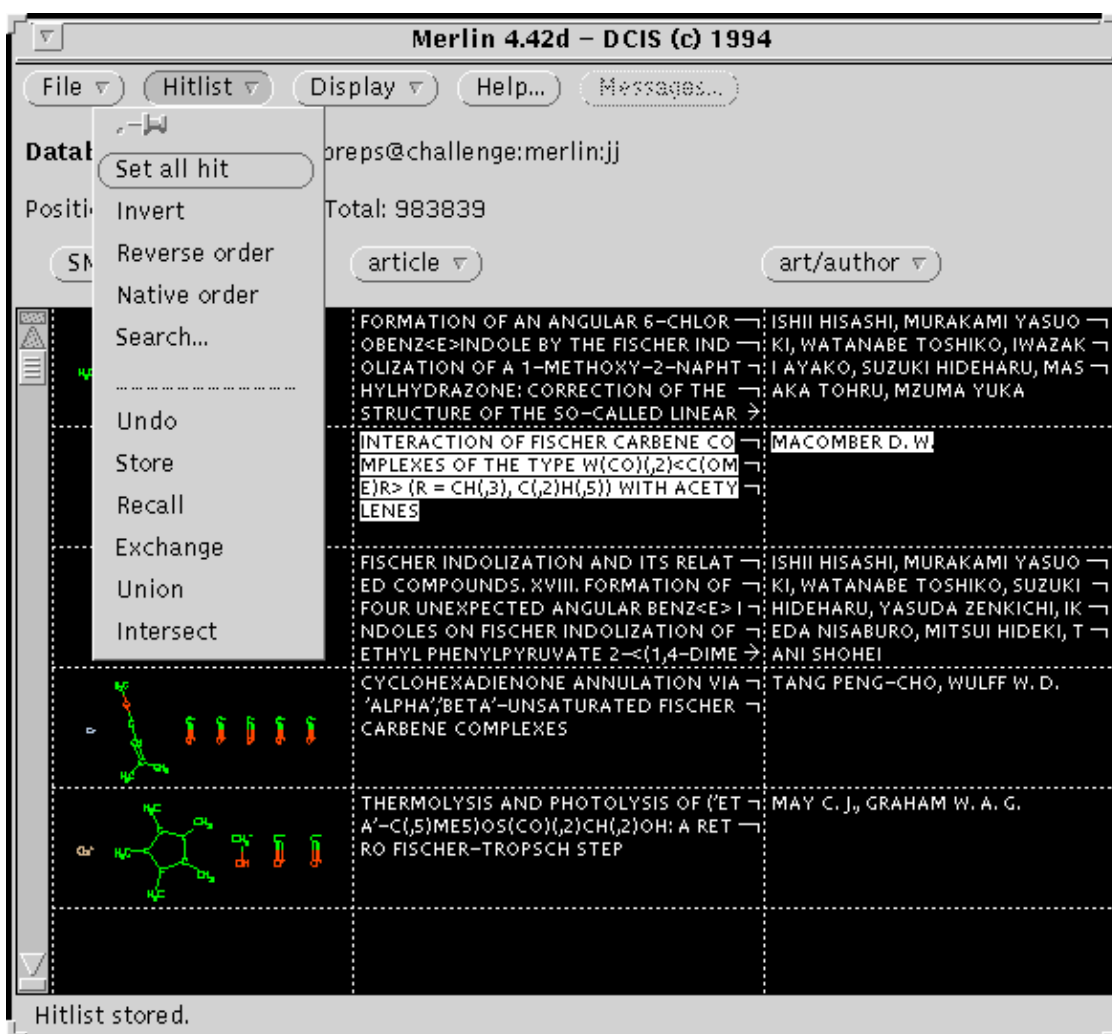
In general, Merlin keeps a data "pool" of structures in memory of which the current "hitlist" is a subset. The hitlist may consist of the entire pool or the null set. The hitlist is a list of "hits" from the previous search, ordered by the previous sort, or in their original pool order. Iterative searching and/or sorting can be used to impose several criteria in a step-by-step fashion, pruning the original pool to a final hitlist of desired structures.

For a non-empty hitlist, there is defined a current structure, indicated by the hitlist pointer (highlighted in the Merlin window). Searches and sorts can be based on this structure and its associated data, although search and sort key data may be entered which is not present in the pool.

The pool is a subset of a THOR database which is loaded into memory by the Merlin server. The hitlist is the subset of those substances which are present as rows on the Merlin scrolling region, visible or not. The set of visible Merlin data columns may or may not include all datatypes present in the pool. Certain functions, such as 'Print hitlist' will print only those datatypes for which columns are visible, though the entire hitlist (including non-visible rows) will be printed. In general, the "hitlist" refers to a set of substances or rows, but not a specific set of associated data. In addition, it is possible to store a hitlist in a buffer and retrieve it, so there is the concept of the "current hitlist" and the stored hitlist. Finally, the function 'Draw hits' is somewhat idiomatic, as it only depicts the hits which are visible.

3. Using the XVMerlin Window Menus

3.1 The Hitlist Menu



The Hitlist menu provides the following functions:

Set all hit

Sets all pool structures "on", so the hitlist contains all the pool

Invert

Structures that are "on" are turned "off", and vice versa

Reverse order

Order reverse with respect to original pool

Native order

Original pool order

Search...

Search window (see sections on searching)

Undo

Undo previous hitlist operation

Store

Store current hitlist in buffer

Recall

Recall hitlist previously stored

Exchange

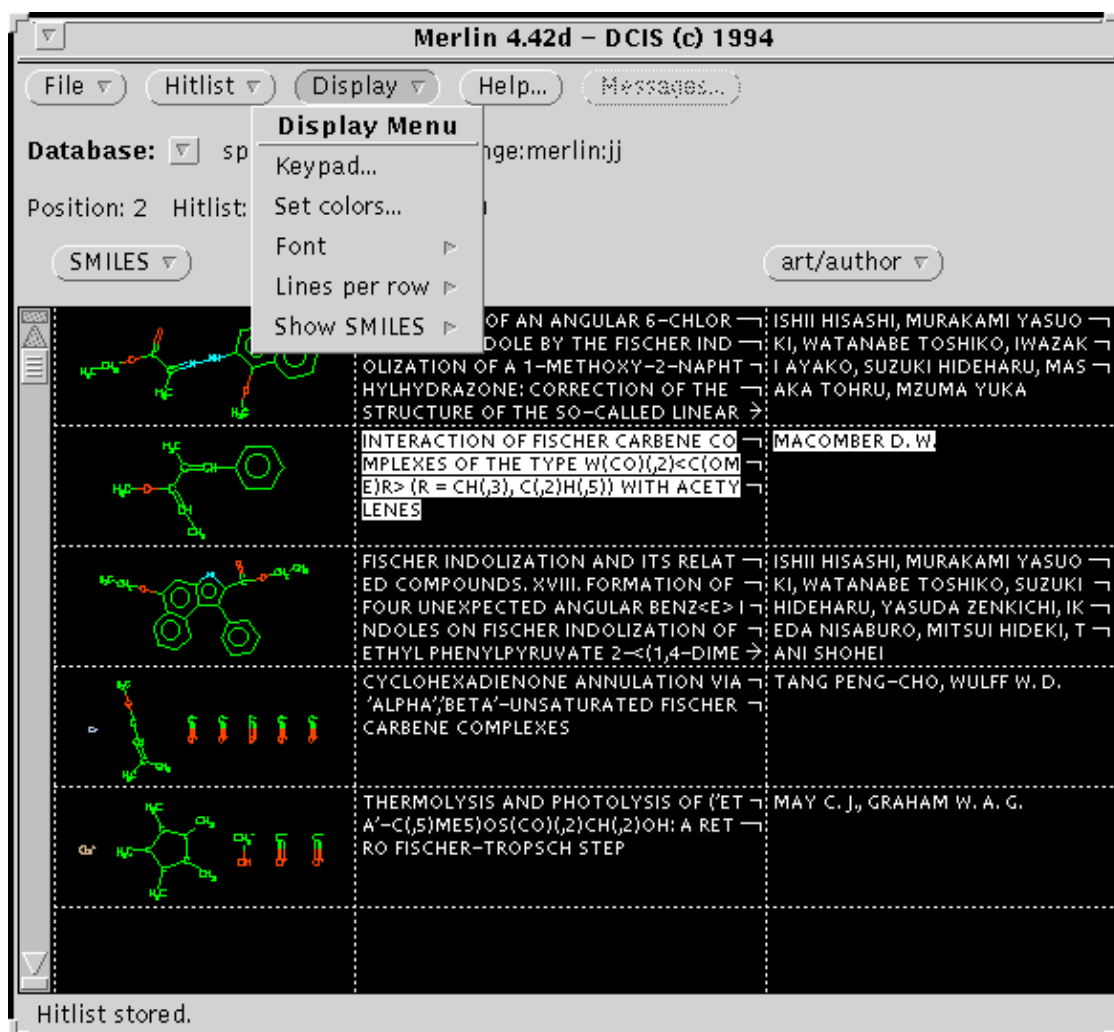
Store + Recall

Union

Union of stored and current hitlist

Intersect

Intersection of stored and current hitlist

3.2 The Display Menu

The Display menu (shown here with 'Lines per row' submenu) provides the following functions:

Keypad...

See below

Set Colors...

Invokes EDGAR, the graphics-attributes widget

Font

Submenu to modify font characteristics

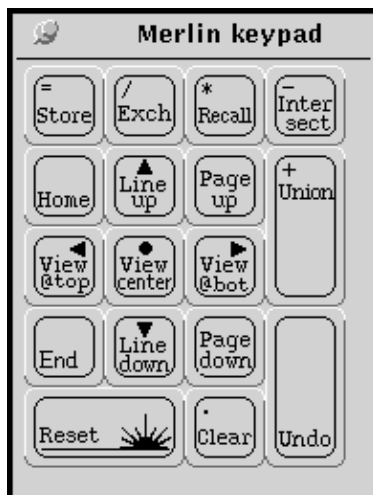
Lines per row

Specifies how many text-lines per hitlist row

Show SMILES

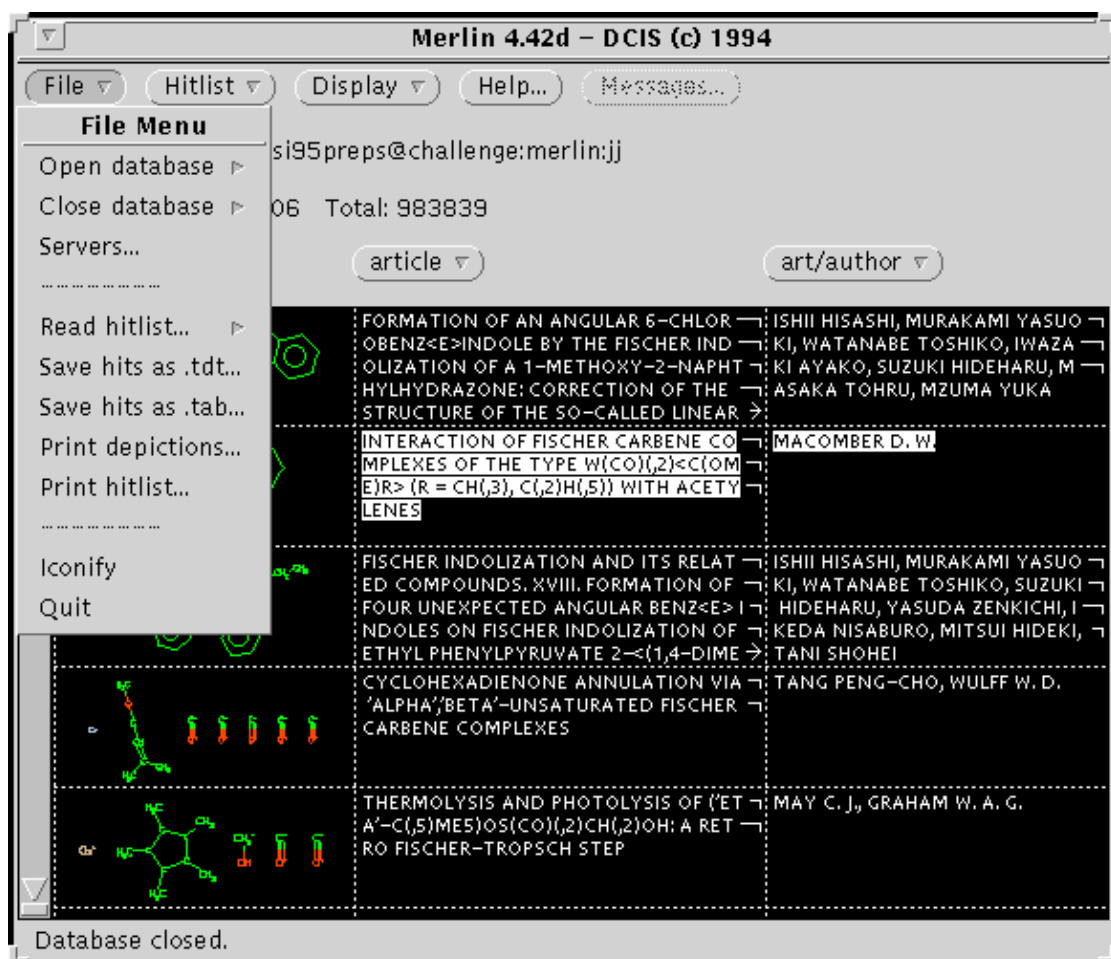
Specifies textual or graphic SMILES (depiction). Note that lines-per-row minimum is 3 for depictions to be displayed.

3.3 The Keypad



The keypad provides handy access to several tools for manipulation of the hitlist, some of which are also present in the hitlist-menu, and identical in function. The Reset button is identical to 'Set all hits'. Home, Line-up, Page-up, View@top, View@center, View@bottom, End, Line down, and Page down all scroll throughout the existing hitlist and/or move the hitlist pointer from one row to another.

3.4. The File Menu



The File menu provides the following functions:

Open Database

View and select from available servers and databases

Close Database

Close any open database

Servers...

Invoke servers control panel (see below)

Read hitlist...

Retrieve hitlist saved as SMILES or other root-ids

Save hits as .tdt...

Save current hitlist

Save hits as .tab...

Save current hitlist in a tab-delimited file

Print depictions...

Print depictions for current hitlist

Print hitlist...

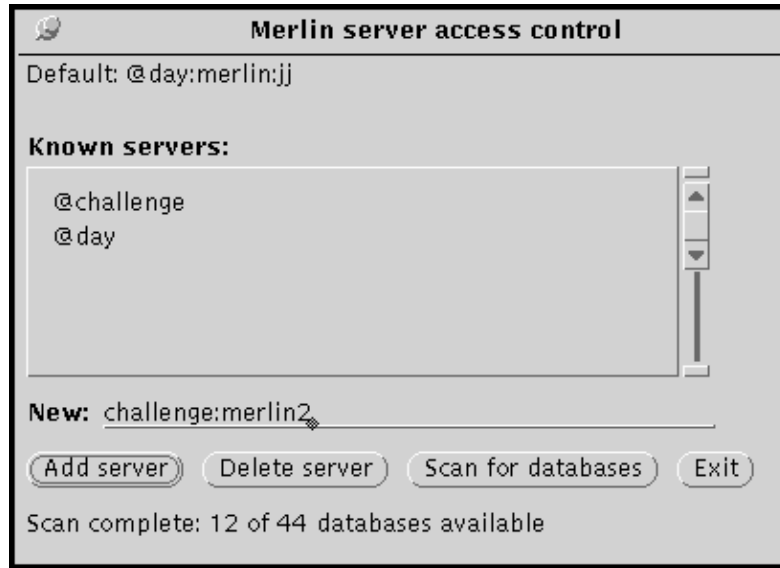
Print current hitlist

Iconify

Iconify XVMerlin

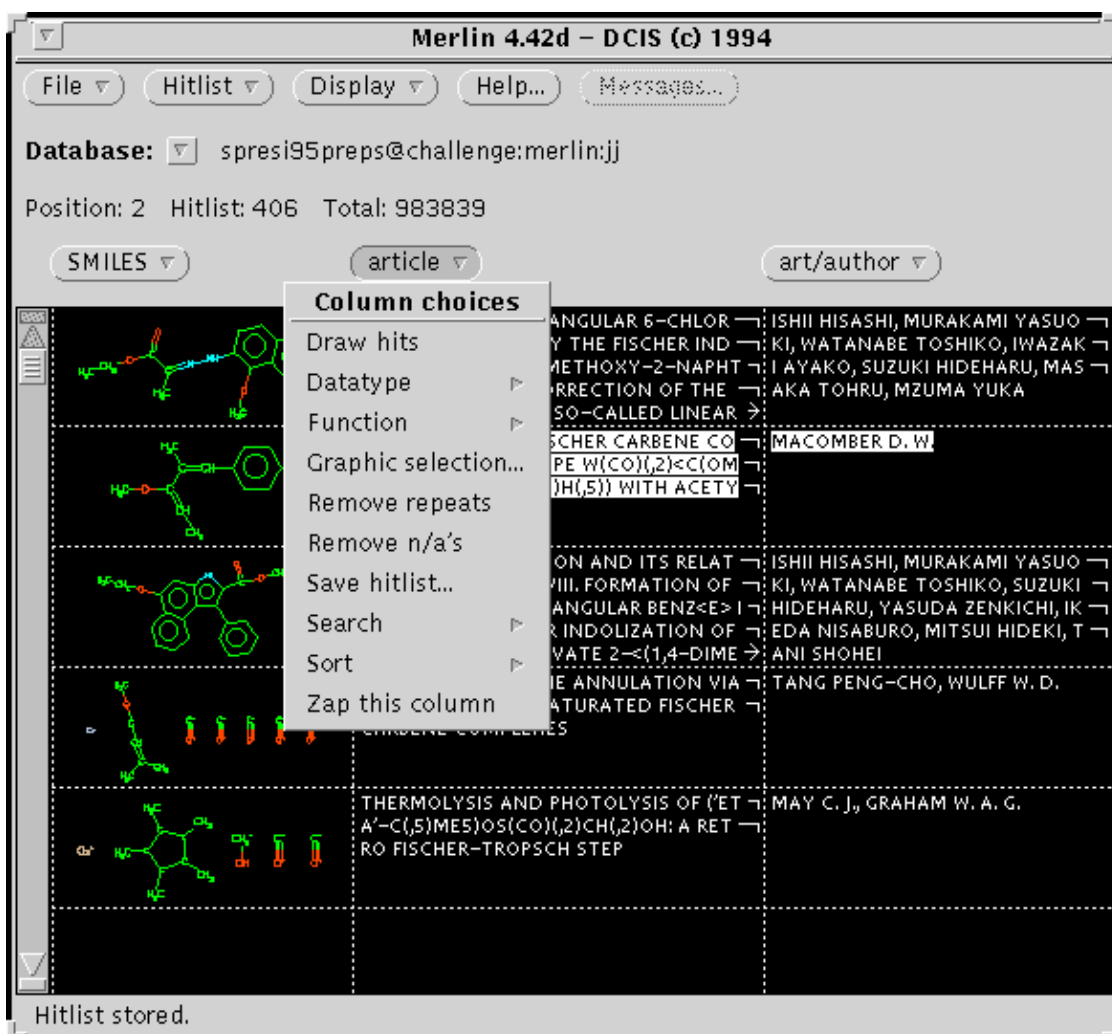
Quit

Quit XVMerlin



The server panel provides access to merlin servers on the network. Merlin will start up with the servers specified by option MERLIN_SERVER_LIST, but additional servers may be added.

3.5 The Data Column Menu



The data Column menu provides the following functions:

Draw hits

Creates a window of depictions for all visible rows

Datatype (submenu)

Modifies the column datatype

Function (submenu)

First, last, min, max, longest, shortest, all, count; select among or operate on multiple data for one datatype

Graphic selection...

Edit hitlist graphically

Remove repeats

Deletes rows where this data value is repeated

Remove n/a's

Deletes rows without this datatype

Save hitlist...

save current hitlist as SMILES file

Search (submenu)

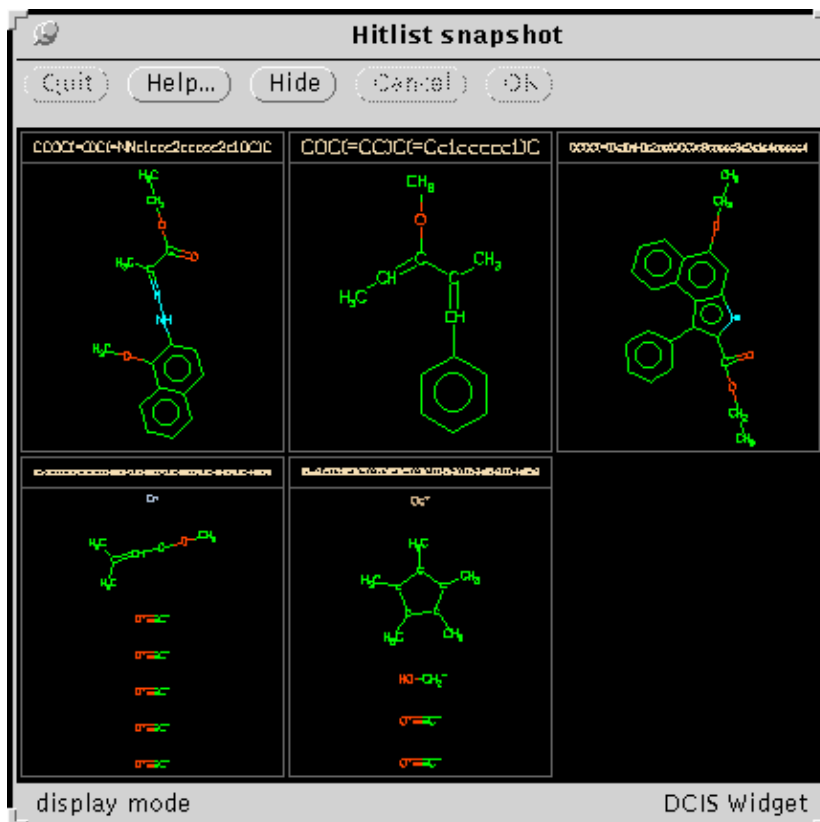
Structural, similarity, and string/expression searching

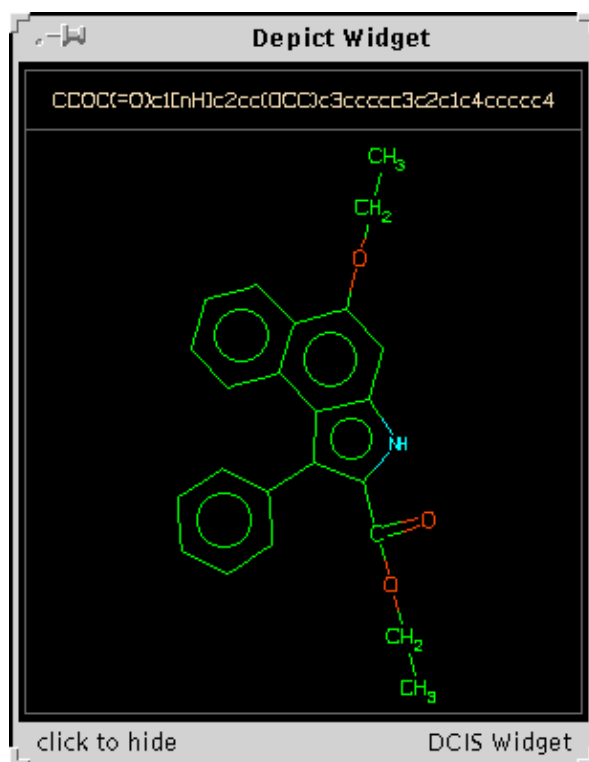
Sort (submenu)

Numeric, ascii, and other sorting

Zap this column

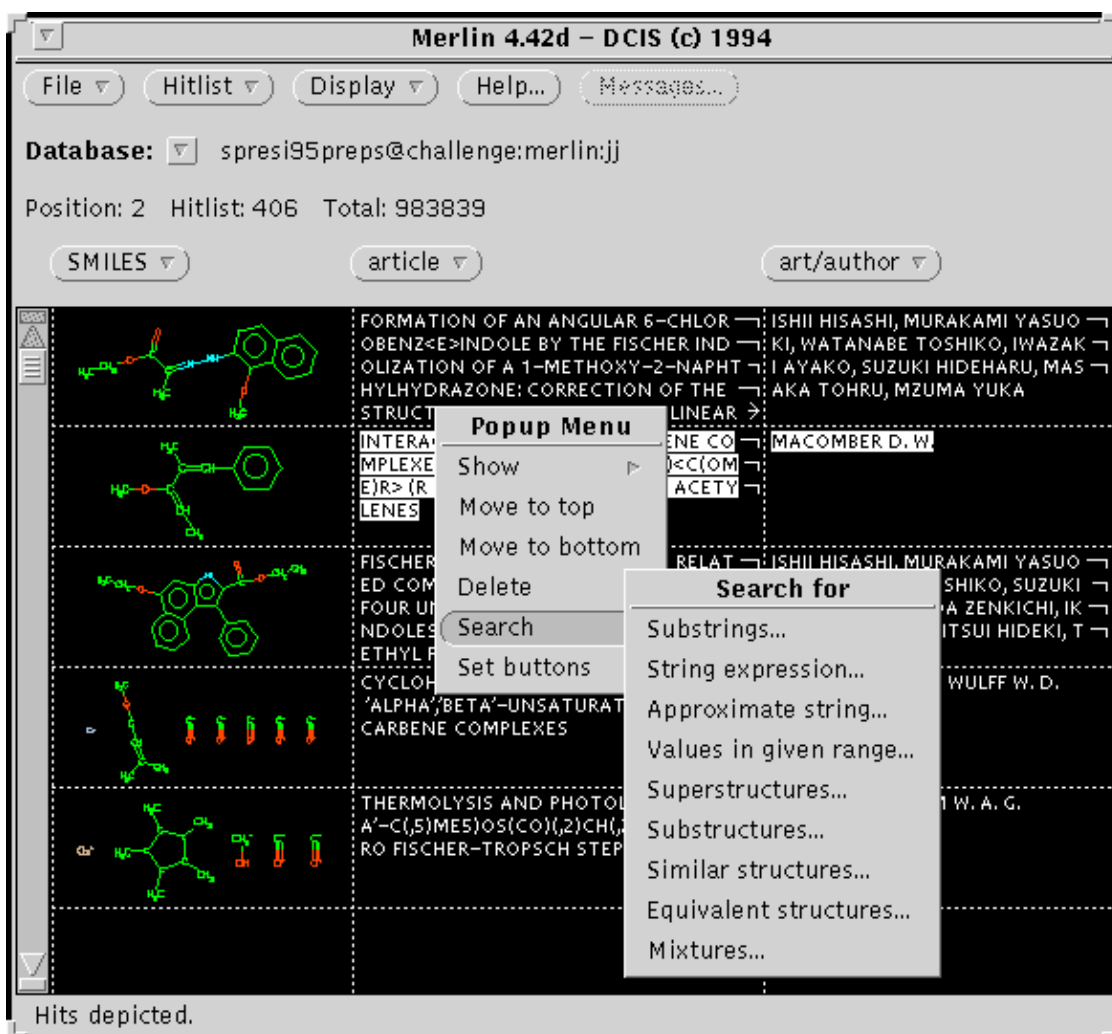
Removes column from scrolling canvas

3.5.1 Draw Hits & the Depict Widget



The Draw Hits command is in each data column menu, and invokes a window depicting all of the structures visible on the hitlist page, subtitled with the datatype of that column. To bring up a larger image of a structure in its own window, use the middle mouse button to click on a pane in the depict widget.

3.6 The Popup Menu



The popup menu is invoked by pressing the right mouse button while the pointer is on a hitlist cell. The popup menu provides functions which involve the row and/or cell from which they were invoked. Often this will save typing in search fields manually.

The Popup menu provides the following functions:

Show (submenu)

Text, 2D or 3D graphics, or the full TDT (via THOR)

Move to top

Moves hitpointer structure to top of hitlist

Move to bottom

move hitpointer structure to bottom of hitlist

Delete (submenu)

Deletes current structure, all above or below, or n/a's

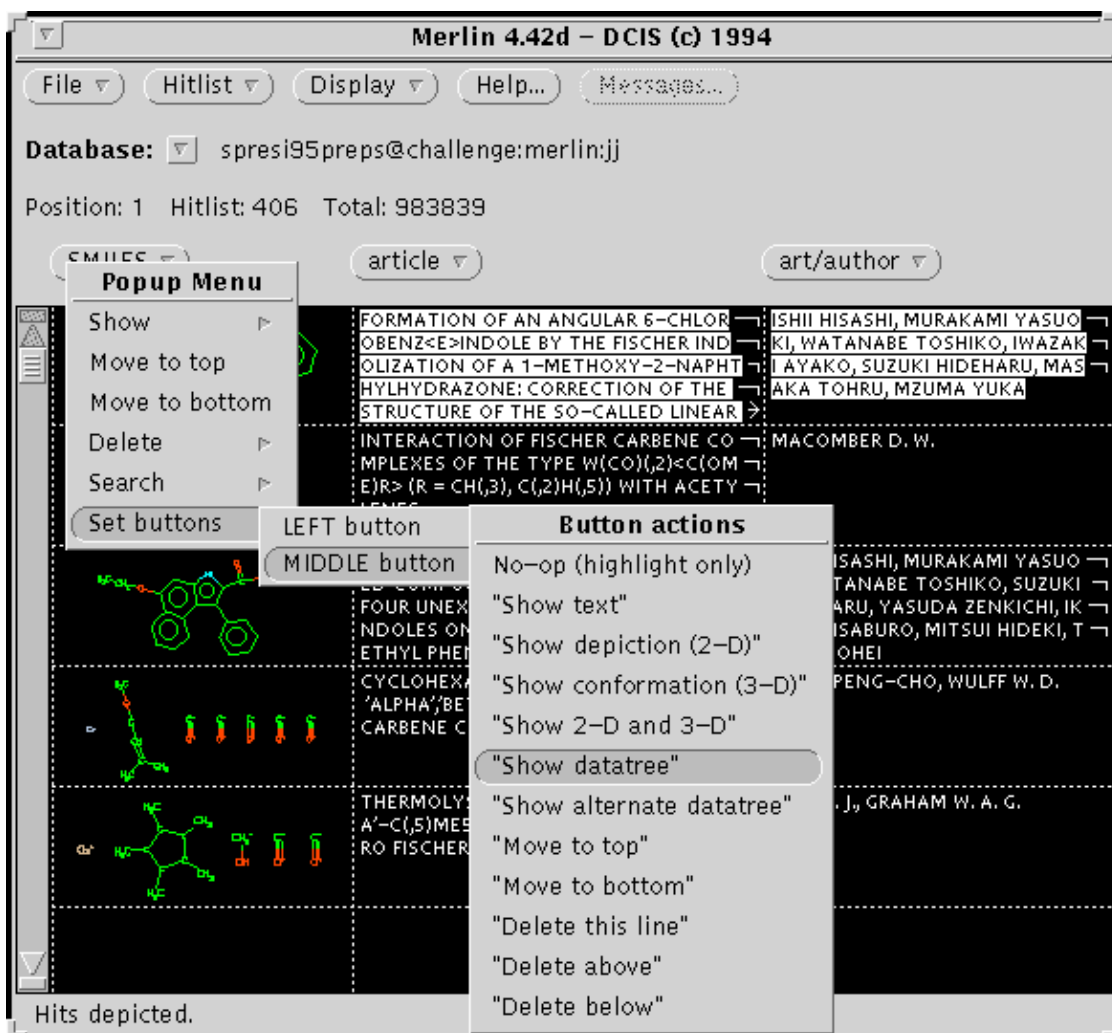
Search (submenu)

Search panel preloaded with cell contents

Set buttons (submenu)

Sets mouse button functions (e.g., show TDT, show 3D)

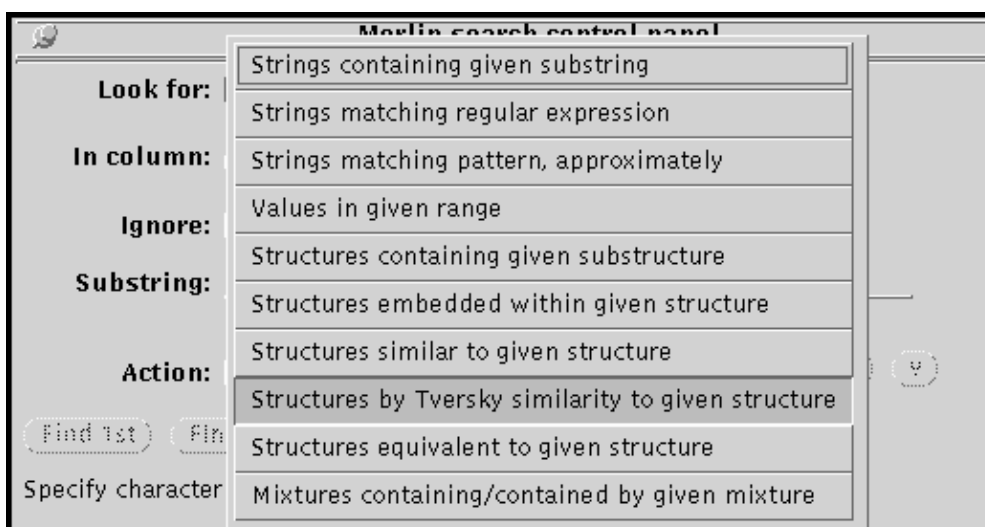
The XVmerlin client can query the THOR server for the entire datatree, or TDT. This capability allows the user to view all data associated with a substance via the handy TDT widget. Use the popup menu from the row of interest to specify Show->datatree. This function is also the default action set for the middle mouse button.



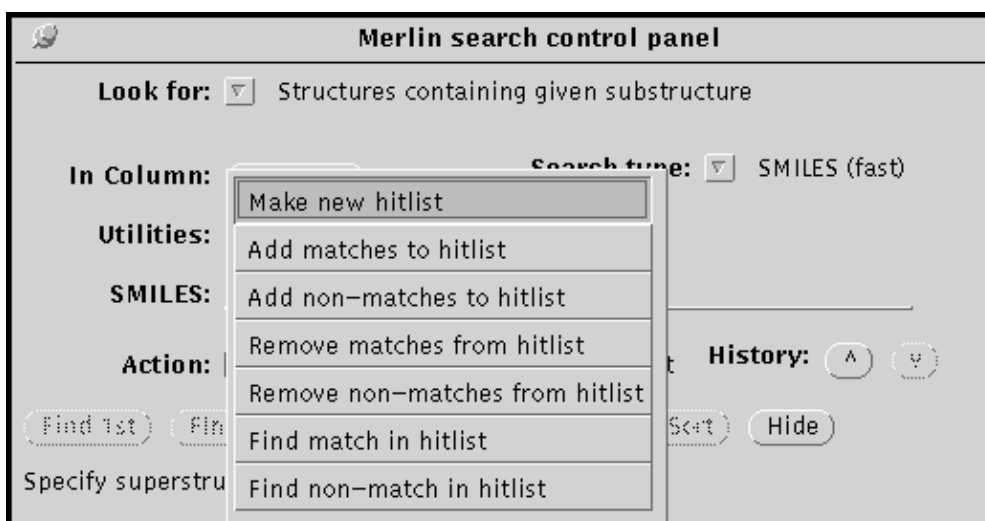
4. Searching

Searching with Merlin means scanning all of a selected datatype (SMILES, name, molecular weight, etc.) for the current hitlist, or the entire pool, or the un-hit portion of the pool, making some comparison or evaluation with respect to a predefined key, and taking a specified action (delete from hitlist, add to hitlist, etc.).

The searching window is invoked from the Hitlist menu, Column menu, or Popup menu. The types of searches are listed in the 'Look for' menu of the Search Control Panel:



And the action taken on the hitlist is specified by the Action menu:



The search-type and action taken are independent choices. So, there is a lot of flexibility over the search procedure, and it is important to be aware of the choices available. In particular, the default action is 'Make a new hitlist', which results in a complete search of the entire pool (hitlist and un-hit). While this is appropriate for a first search, to then search only the hitlist resulting from the first search will require a different action, such as 'Remove non-matches from hitlist'.

The 'Find first' and 'Find next' buttons apply only for the actions 'Find match...' and 'Find non-match...'.

The Search window appears differently for each search type. The search types are:

String search:

Looks for the specified string in the specified column. 'Select & sort' also performs an ascii sort.

Regular expression search:

Looks for the specified regular expression (UNIX-style) in the specified column. Refer to UNIX documentation or a local guru for help with regular expressions.

Approximate-string search.

Ranks according to similarity to given string or regular-expression.

Structures containing given substructure:

In this search the specified structure is searched for as a substructure. GRINS may be used for structure input. The search type may be for SMILES, Isomeric SMILES, SMARTS, or Isomeric SMARTS. SMARTS searches can be much more chemically meaningful, but the search algorithm is much slower. The 'Optimize target'checkbox invokes a routine for rearranging the search target to place uncommon atoms first, to speed up the search. This box should normally be checked.

In general terms, a SMILES search looks for the non-hydrogen graph of the specified SMILES. The SMILES must represent a valid molecule. SMARTS represent substructures and may or may not be valid SMILES. Press the Help button to find documentation on SMILES and SMARTS. Refer also to the Daylight Theory Manual.

Merlin utilizes fingerprints for fast screening of structures as a first step. If the database contains FPP part N-tuple fingerprints, this screen can be optimized by checking the "Use FPP" box.

Structures embedded within given structure:

This search looks for SMILES which represent substructures of the specified structure.

Similarity search:

Compares the fingerprint for the specified SMILES with the fingerprints in the pool. For each comparison a similarity coefficient is generated (and this coefficient can be displayed as a column of datatype "Similarity"). This coefficient is generated by the Tanimoto similarity algorithm. Merlin can either sort the hitlist on this coefficient, or select and sort, deleting those SMILES whose coefficient is less than an arbitrary value. The "very high" through "rough" choices represent these arbitrary numerical thresholds, and these values can be reset by options.

The correspondence between Merlin-similarity and chemical similarity will, of course, be dependent on the user's definition of chemical similarity. The fingerprint program is equipped with adjustable information-content and information-density settings, which can improve similarity searching for any given dataset. It is important to recognize that no similarity metric is likely to be optimal for all chemical tasks.

Tversky similarity search:

The most powerful structural search is now (as of 4.51) the Tversky search (but not as simple to use or interpret as the Tanimoto metric). Like the Tanimoto search, this compares features in a given structure (the "prototype") to features in database structures (as "variants"), and allows hitlist selection or sorting based on the results. However, the Tversky search allows you to specify the weighting that will be given to each set of features.

Setting the weighting of prototype features to 100% and variant features to 100% produces a symmetrical similarity metric identical to Tanimoto metric. (Setting them symmetrically to values less than 100% doesn't change the rank ordering, just the absolute value, i.e., more structures will meet a

given similarity criterion).

Setting the weighting of prototype and variant features asymmetrically produces a similarity metric in a more-substructural or more-superstructural sense. Setting the weighting of prototype features to 100% and variant features to 0% means that only the prototype features are important, i.e., this produces a "superstructure-likeness" metric. In this case, a Tversky similarity value of 1.0 means that all prototype features are represented in the variant, 0.0 that none are. Conversely, setting the weights to 0% prototype / 100% variant produces a "substructure-likeness" metric, where completely embedded structures have a 1.0 value and "near-substructures" have values near 1.0. (Note: with no weight at all given to variant features, this metric is pretty sensitive fingerprint "noise" and settings of 90%/10% generally produce a more useful ranking.)

Tversky metrics where the two weightings add up to 100% (1.0) are of special interest (e.g., the 50/50 metric is known as the Dice index). The Tversky search query panel provides a "Sum 100%" checkbox which, when selected, forces the two weights to add up to 100%.

Advanced users may wish to experiment with Tversky metrics where weightings are not limited to 100%. (Doing so is rank-equivalent to raising the Tversky "theta" parameter above 1.0.) Weightings greater than 100% causes the distinguishing features to be emphasized more than common features which may be useful in analysis of diversity or dissimilarity. The Tversky query panel does not provide control of the maximum allowed weighting directly; it must be set with an option, e.g., `xvmerlin -MERLIN_TVERSKY_ABMAX 2.0`

Four xvmerlin options are used to control the default Tversky parameters: `MERLIN_TVERSKY_ALPHA` (prototype weighting), `MERLIN_TVERSKY_BETA` (variant weighting), `MERLIN_TVERSKY_ABMAX` (maximum weighting), and `MERLIN_TVERSKY_ONE` (setting of "Sum 100%" checkbox).

Equivalent search:

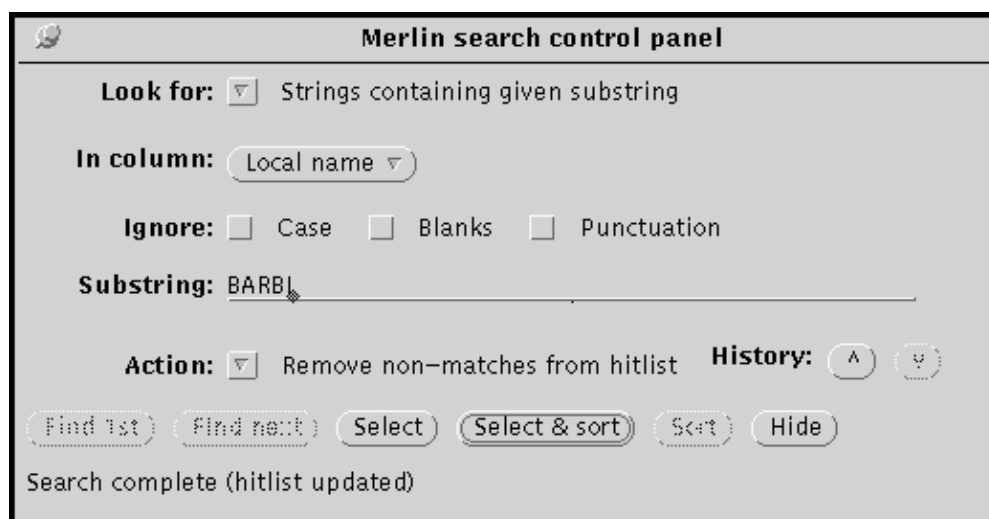
This search has three distinct modes. An equivalent-SMILES search is the same as a Thor lookup for the given SMILES. A graph search looks for all structures with the same hydrogen-stripped, bond-order-stripped graph. A tautomer search looks for graph-equivalents with the same formula and charge.

4.1. Sample Searches

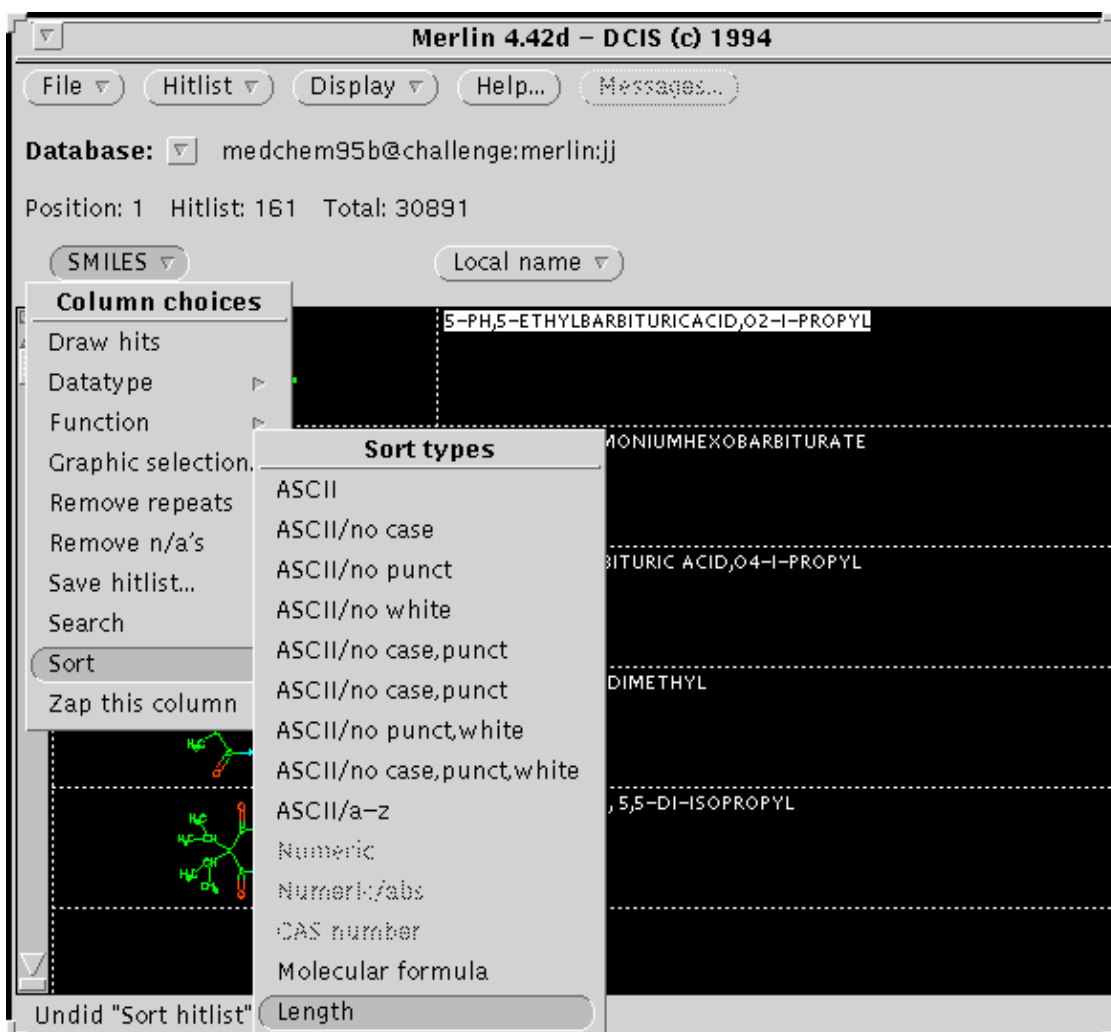
The first search described is a structural search; the second is a string search. Note that the flexibility of Merlin allows many options in performing these searches. These are only two possible procedures.

4.1.1 A String Search

1. From the XVMerlin window, invoke the Search Command Panel by choosing 'Searching' from the Hitlist menu.
2. From the 'Look for' menu, choose 'Strings containing given substring'
3. From the 'In column' menu, choose datatype 'Local name'.
4. Enter the string "BARBI" and press Select.



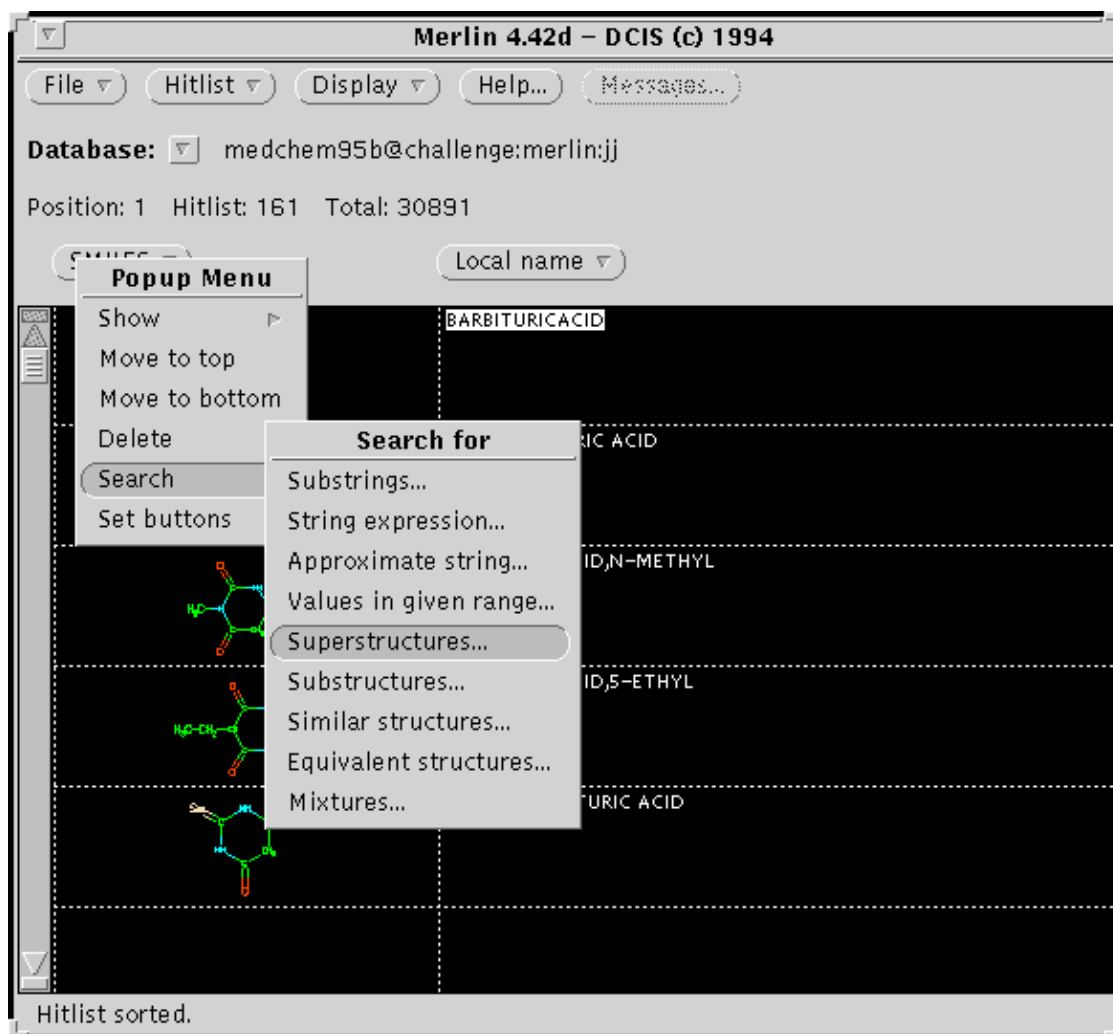
5. Use the popup menu from the SMILES column to sort the hitlist based on SMILES length.



- Use the scroll bar (or the keypad) to find the first SMILES.

4.1.2 A Superstructure Search (SMILES)

- From the previous string search and sort, Barbituric Acid should be at the top of the hitlist. Use the scrollbar to make the first row the current structure.



- Invoke the popup menu from Barbituric Acid. Select the Search panel in 'Superstructure search' mode. Note that the Search panel is preloaded with the corresponding SMILES. Press the Select & sort button to start the search. Sorting will be based on similarity. After the search, note that depictions are highlighted to show the hit atoms.

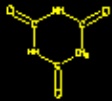
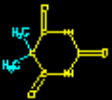
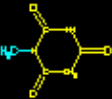
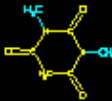
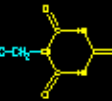
Merlin 4.42d - DCIS (c) 1994

File ▾ Hitlist ▾ Display ▾ Help... Messages...

Database: ▾ medchem95b@challenge:merlin:jj

Position: 1 Hitlist: 123 Total: 30891

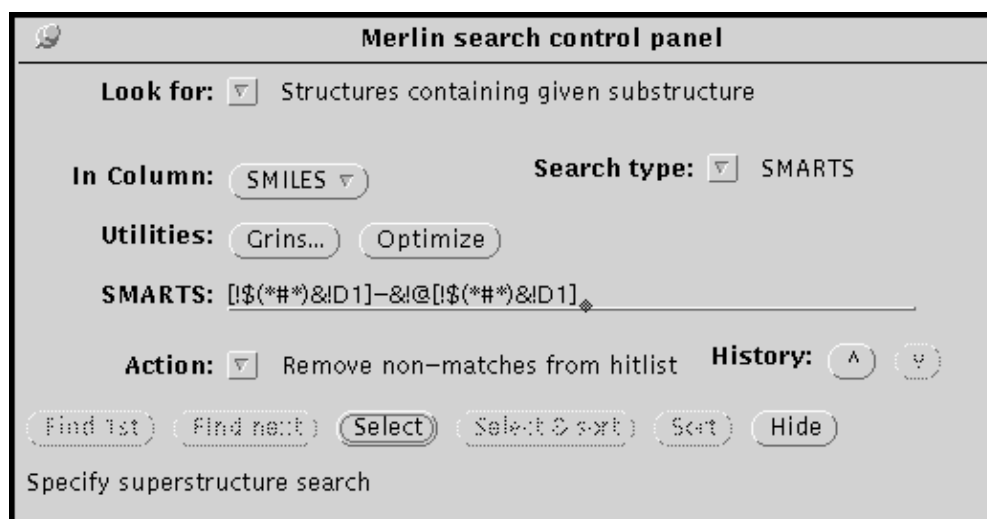
SMILES ▾ Local name ▾ similar ▾

	BARBITURICACID	1
	BARBITURICACID,DIMETHYL	0.9545
	BARBITURICACID,N-METHYL	0.9114
	BARBITURIC ACID-1,3-DIMETHYL	0.9114
	BARBITURICACID,5-ETHYL	0.8889

New column added to hitlist display.

4.1.3 A SMARTS Search

1. SMARTS searching is slower than SMILES searching, but is more powerful. With the Search panel in Superstructure mode, specify SMARTS searching, and type in the following SMARTS: [!\$(***)&!D1]-&!@[!\$(***)&!D1]. Then press Select. Note that this SMARTS represents two atoms connected by a "rotatable bond".



2. This search may take a few minutes. The status widget will report progress.
3. The resulting depictions are highlighted to show the hit atoms.

The screenshot shows the Merlin 4.42d interface with the following details:

- Window Title: Merlin 4.42d - DCIS (c) 1994
- Menu Bar: File, Hitlist, Display, Help..., Messages...
- Database: medchem95b@challenge:merlin:jj
- Position: 1 Hitlist: 27387 Total: 30891
- Buttons: SMILES, Local name, similar
- Table of Results:

Chemical Structure	Local name	Similarity
	TRIMETREXATE	
	EPROSARTAN	~
	2,4,6-TRIIODOPHENOL,TETRAETHYLAMMONIUMSALT	~
	GAPROMIDINE	~
	BUFURALOL	~
- Status Bar: Search complete (hitlist updated)

5. Configuring Merlin

There are several steps that can be taken to configure Merlin so that databases are loaded and opened automatically and displayed to the user in a convenient and preferred way. These steps can be divided into two categories, Merlinserver configuration issues, which are not covered in this manual, and xvmerlin-user configuration issues, which are. For clarity, these are some Merlinserver configuration issues not covered here but important to the overall configuration:

- Starting the Merlinserver and Thorsserver (maybe automatically)
- Loading specified databases
- Loading specified datatypes and datafields
- Server and database security

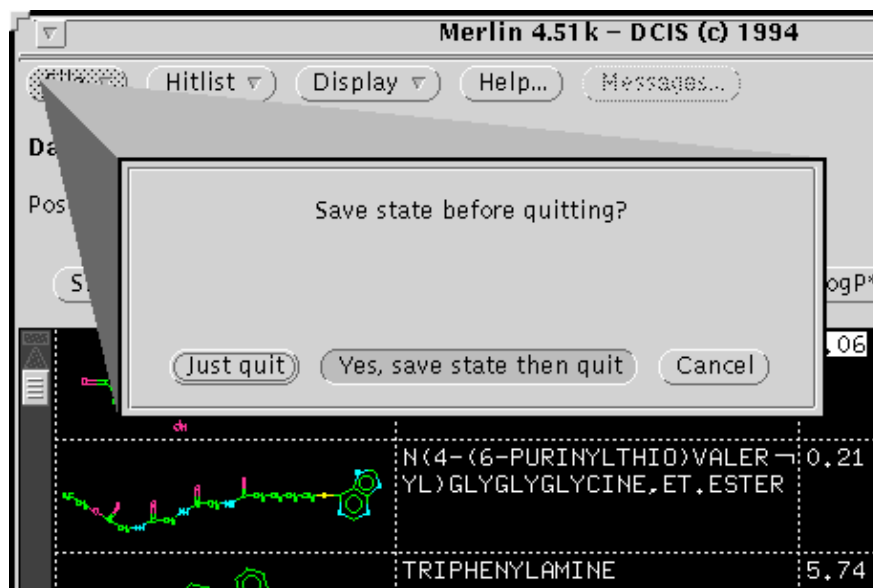
In the user's environment, there are a few Daylight options which affect the Merlin environment. Foremost among these is `MERLIN_SERVER_LIST`, which should be set to a comma-separated list of Merlin servers. This can be done in the user's Daylight profile file, or by environment variable, i.e., in `csh`:

```
setenv DY_MERLIN_SERVER_LIST
"challenge,corona,cojones:merlin:thor"
```

Saving and restoring XVMerlin's state

After opening databases and defining columns with xvmerlin, you may save the state of the program so that the same configuration is restored when restarted. To accomplish this, follow these steps:

1. Set up xvmerlin as desired.
2. Quit using the File->Quit command. Press the "save state" button:



3. Provide a password for reopening xvmerlin when prompted.



4. The previous actions will result in an xvmerlin configuration file being saved in \$HOME/.dy_merlinprofile.opt. However, to invoke this file it must be included in the user's Daylight profile file, by default \$HOME/dy_profile.opt or respecified by environment variable DY_PROFILE. Add the following line to your profile file:

```
#include $HOME/.dy_merlinprofile.opt
```


Daylight XVMerlin Manual

Now restart xvmerlin. You should be prompted for a password, and the original saved configuration should be restored.