PCModels Manual

Table of Contents

PCmodels Manual

A User Guide for the XView user interface to clogp and cmr-xvpcmodels

Daylight Version 4.9
Release Date 08/01/11

Copyright notice

1. About PCmodels; Scope of this manual

PCmodels is a Daylight software module providing access to two chemical models: CLOGP(hydrophobicity partition coefficient) and CMR(molar refractivity). These models are developed and supported by BioByte, Inc. of Claremont, CA. Daylight has worked with BioByte to offer PCmodels in its current form: a convenient user-interface compatible with Daylight's other software. A window interface is provided for interactive access(**xvpcmodels**), command line filters enable batch processing(**clogp** and **cmr**), and an object-oriented

interface is available providing high-performance access to external programs (**clogptalk** and **cmrtalk**). Documentation for PCmodels consists of several parts:

- This "User Guide", intended to provide a working knowledge of interactive and batch interfaces to the PCmodels algorithms.
- The "CLOGP Reference Manual" which describes the theory and methodology of the CLOGP algorithm for hydrophobicity estimation, including examples.
- The "CMR Reference Manual" which describes the theory and methodology of the CMR algorithm for polarizability estimation, including examples.

2. Getting Started

Prerequisites for running the interactive PCmodels program xvpcmodels:

- The PCmodels module has been installed (see installation instructions)
- Local environment variables have been defined.
- The Daylight software license is valid for "**xvpcmodels**", "**clogp**", and "**cmr**".
- To start the program enter "**xvpcmodels**" (on IRIX/SGI, enter "**xvpcmodels4d**").

See Section 4 for batch processing environment.

3. Interactive PCmodels usage: xvpcmodels

The **xvpcmodels** program provides interactive access to **pcmodels** for users in the X-Windows environment (the "xv" prefix stands for "XView"). The window initially displayed when **xvpcmodels** starts up consists of several pull-down menus and buttons, and two "text panes". Structures must be specified as SMILES in the input text pane. SMILES may be entered graphically via GRINS, lexically by typing directly into the input text pane, or by loading a file containing a SMILES into the input text pane. Model selection and output format are specified by pull-down buttons. Pressing the 'Compute' button runs the model; output is displayed in the output text pane.
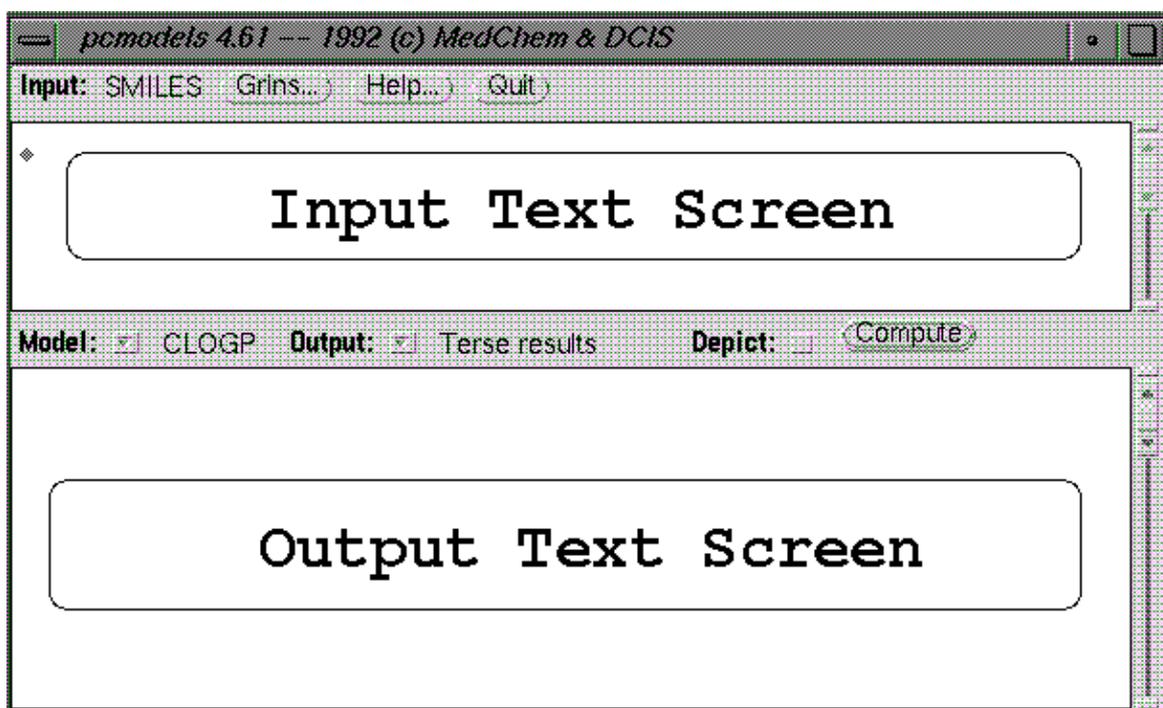
3.1. Defining an input structure

The simplest way to use **xvpcmodels** is to locate the mouse in the input text pane, type a SMILES, and click the Select(left) mouse on the Compute button. (If you are unfamiliar with SMILES, enter "C", the SMILES for methane.) Output will appear in the output text window. If the CLOGP value is always returned 0.00, the Daylight option PCMODELS_FRADB_BIN is currently undefined or invalid.

GRINS provides graphical structure specification. To use GRINS, press the Grins button and the GRINS window should appear, as shown below:
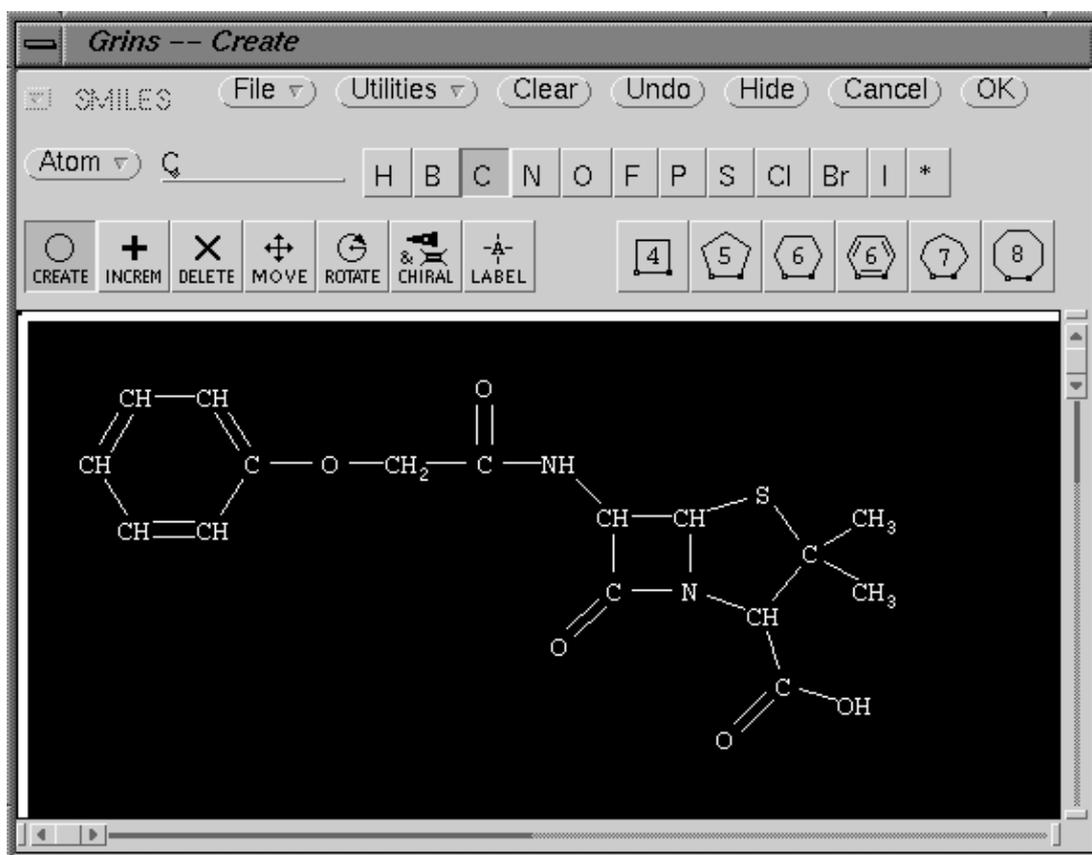
The GRINS widget is described in the *Daylight Widgets User Guide*. Refer to that document for detailed instructions. GRINS allows the user to graphically edit a structure from atoms, bonds, templates, and parent molecules. When the structure has been selected in GRINS, the SMILES for the structure will appear in **xvpcmodels'** input text pane.

A third way to define the input structure is to load a SMILES file into the input text pane, locate the cursor on the desired SMILES, and then press the Compute button. Note that the triangular cursor in the input text pane may be placed anywhere on the SMILES line; the syntax of a SMILES line is a SMILES string followed by an optional space and name.
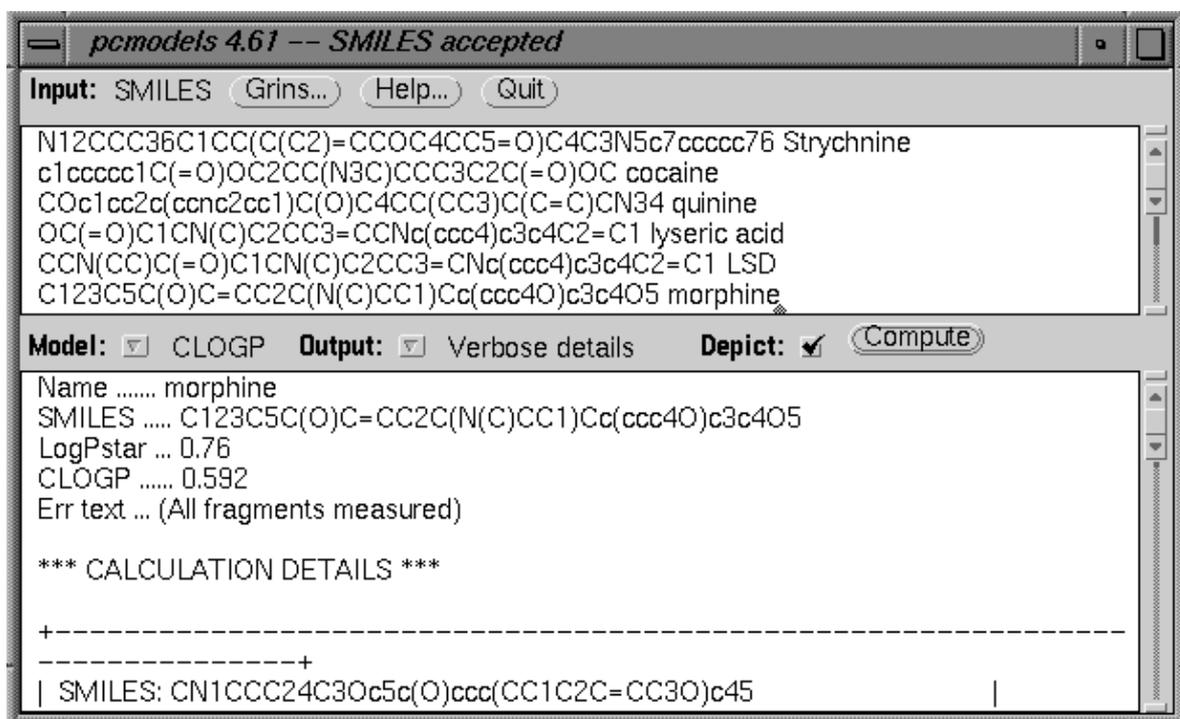
Refer to the section "Text operations" in the following pages for details of loading text files, writing text files, and handling text files using the text pane windows and popup menus.

3.2. Selecting Model CLOGP or CMR

Choose CLOGP or CMR by using the "Model" pulldown menu, as shown below:
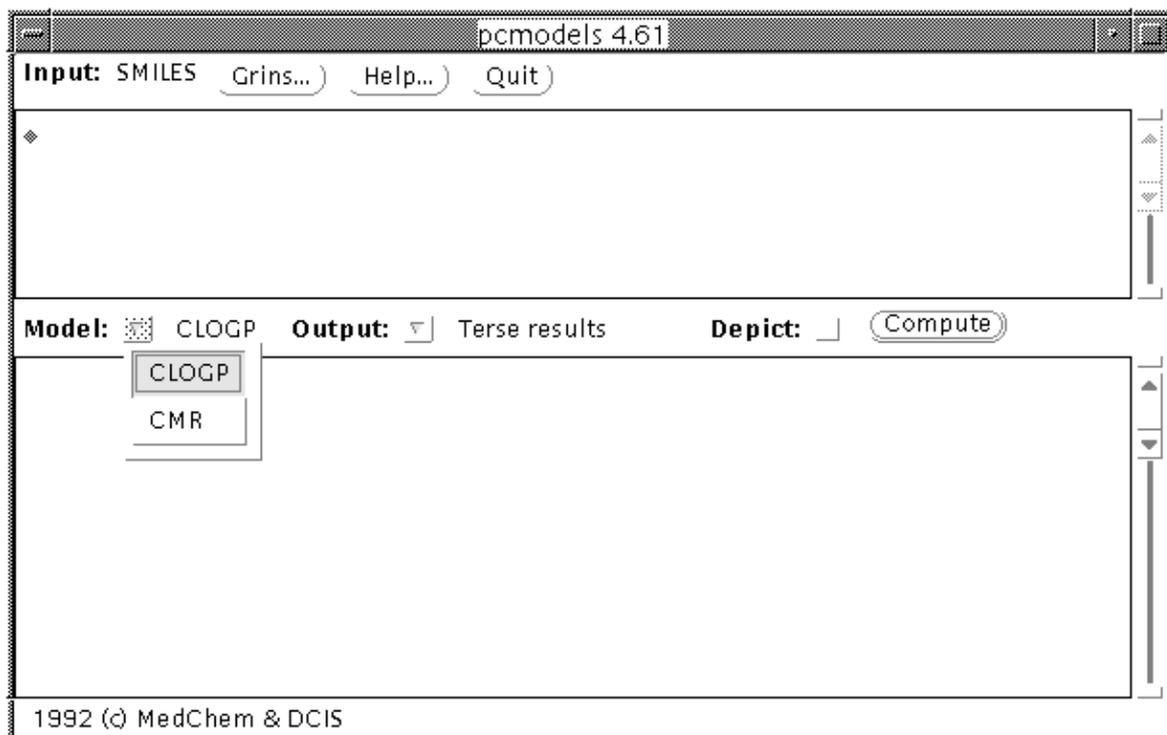
3.3. Selecting Output Level

The "Output" pulldown menu allows the user to specify output detail level:
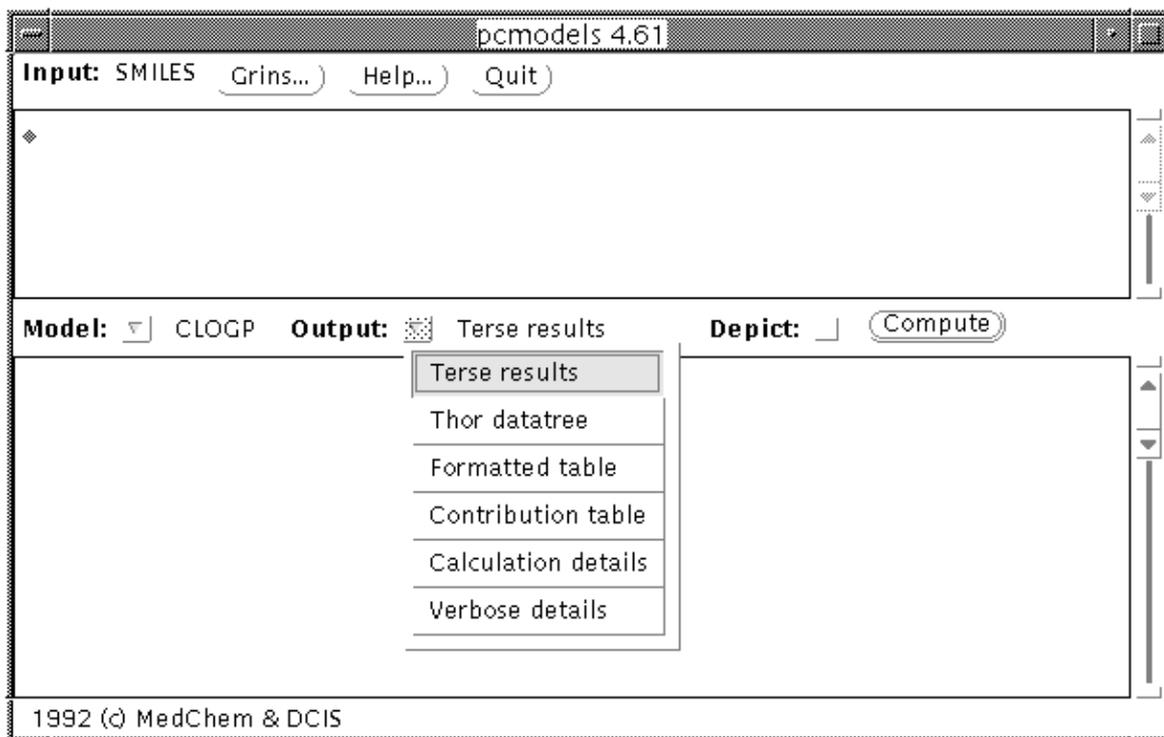
Changing the output level invokes model computation.

3.4 Text Operations

The input and output text panes and the text pane popup menu are standard XView tools, and are very versatile. Refer to the XView documentation and Help-key documentation for a complete description. A few of the capabilities are shown below. The text popup menu can be invoked for either the input text pane or the output text pane. Just locate the mouse cursor in the desired pane, and press the mouse menu button (right button). Then drag to pop up a submenu.
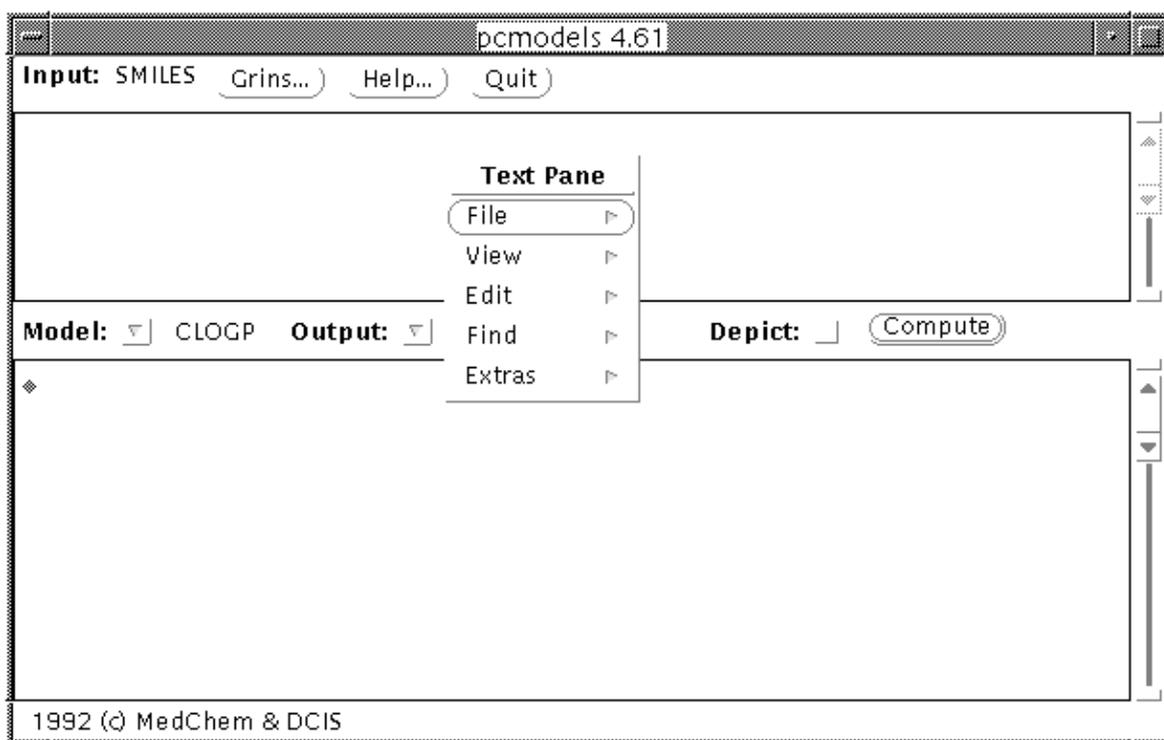
To load a file into the input text pane, select "Load file"/"Open" from the Text-File submenu. To save the contents of the output text pane, select "Store as New File"/"Save As" from the Text-File submenu.
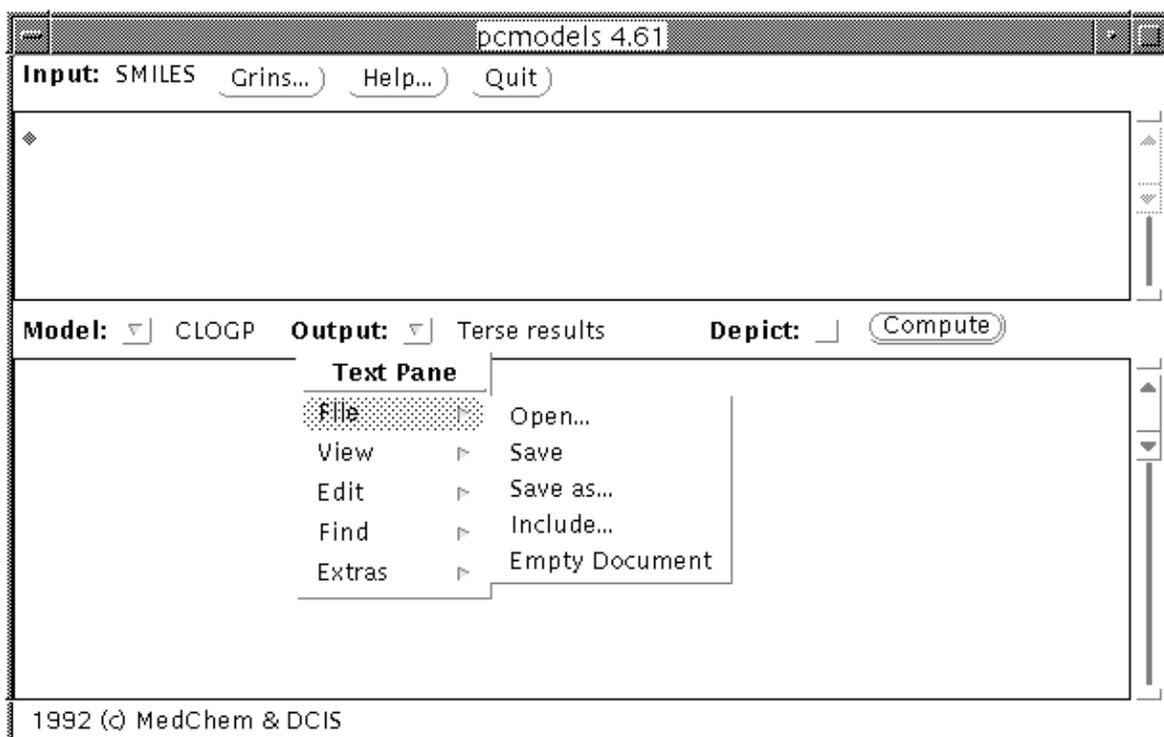
To find a selected string, select "Find Selection" from the Text-Find submenu. To find the same string again, select "Again" from the Text-Edit submenu.
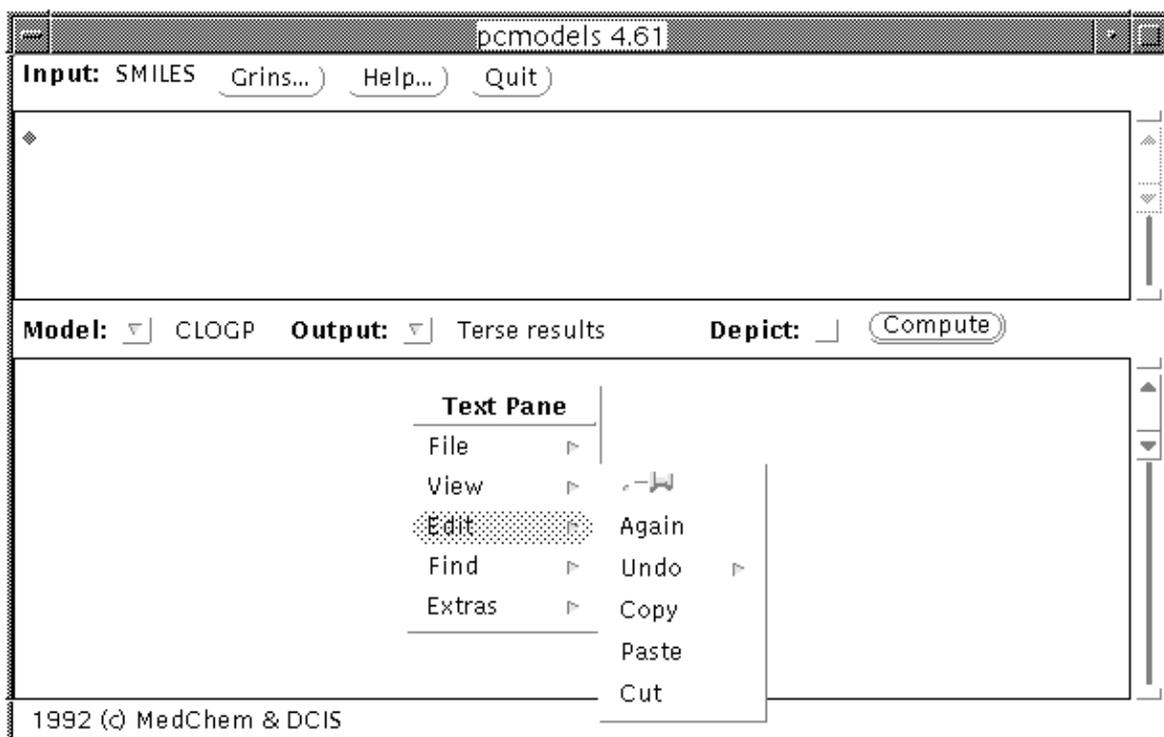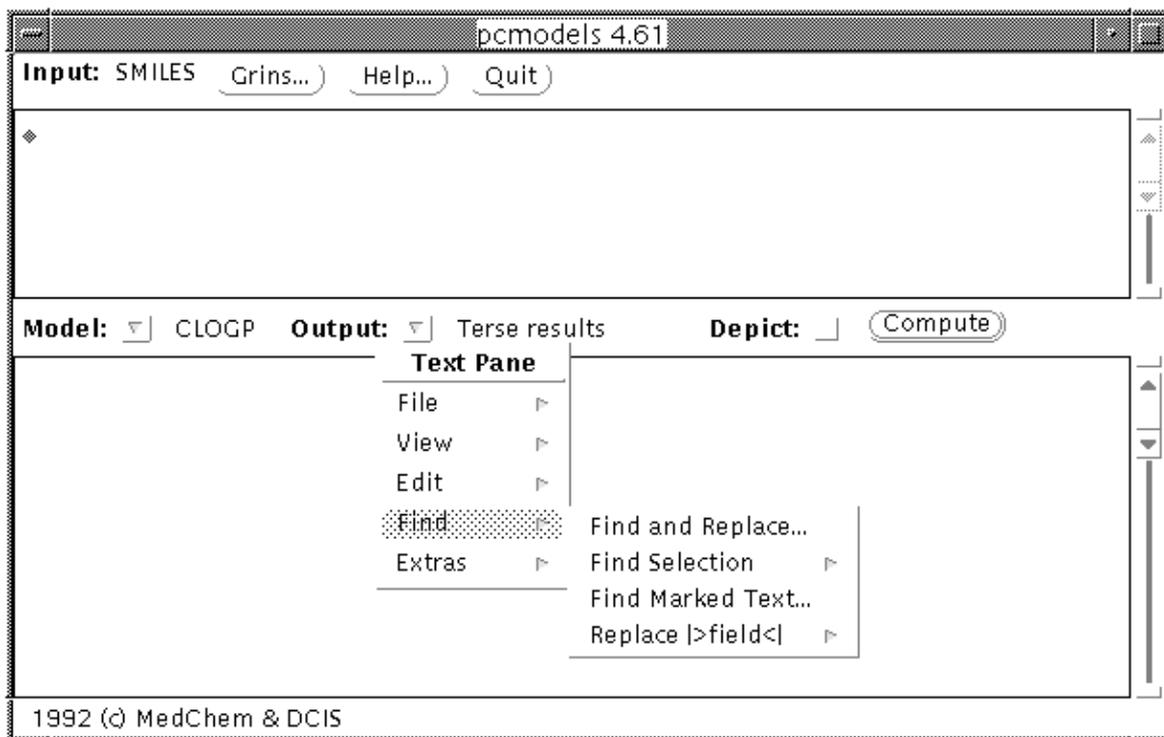
3.5. Depict, Help, etc.

The Depict widget for 2D structure display can be invoked by selecting the Depict box on the main
xvpcmodels window. Depict will appear as a separate window with 2D depiction of the selected structure and
the result of the selected model. If the selected model is CLOGP, CLOGP fragments will be highlighted.
Checking the "Depict" box automatically invokes model computation.

Depict Widget -- display mode

Qui Help... Hide Cancel OK

Cl

N

H₂C

C —— NH

H₂C

NH

Cl
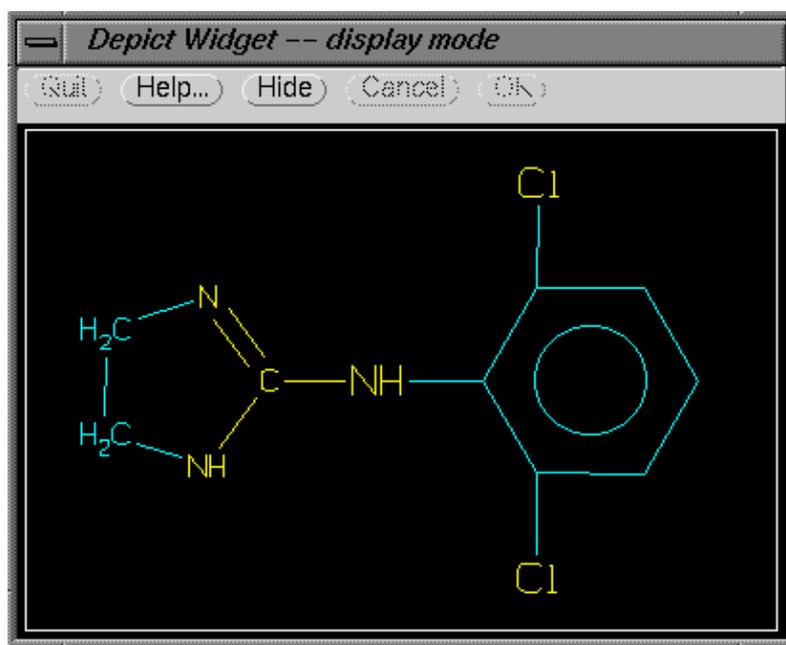
The Help widget is invoked by selection the Help button on the main **xvpcmodels** window.

4. Using PCmodels batch filters

The programs provided for batch processing of structure files are "**clogp**" and "**cmr**".

These programs read in a TDT(THOR Data Tree) file and write out a TDT file with CLOGP or CMR data added. The program **clogp** is used in the following examples; **cmr** has identical commands.

To add CLOGP data to a TDT file, enter:

% **clogp < infile.tdt > outfile.tdt**

or

% **clogp infile.tdt outfile.tdt**

If no output file is specified, output is sent to the screen. To write a help message to the screen type:

% **clogp –HELP**

```
 +----------------------------------------------------------+
| CLOGP -- Add CLOGP results from/to .TDT files            |
+----------------------------------------------------------+
| Copyright  1990-2011 (c) Daylight CIS, Inc.              |
| Copyright  1985-2011 (c) Pomona College                 |
+----------------------------------------------------------+
| $ clogp [options] [infile.tdt [outfile.tdt]]            |
|     Copy .TDT input file to output file, adding CLOGP    |
|     polarizability estimate as a CLOGP datum.  The CLOGP |
|     datatype has 3 fields:  CP.    |
|     Output to standard output if second argument is      |
|     missing. Input from standard input if no arguments   |
|     are given.  Exit w/ warning if CLOGP is inaccessible;|
|     copy unchanged input tree to output on other errors, |
|     e.g., invalid input, CLOGP can't make valid estimate.|
```

```
|                                                          |
| $ clogp -HELP [TRUE|FALSE]                               |
|      Write this help message to standard output and exit. |
|                                                          |
| $ clogp -PCMODELS_VERSION [TRUE|FALSE]                   |
|      If true, writes the version numbers of the program, |
|      Biobyte algorithm, and database to standard output. |
|                                                          |
| $ clogp -PCMODELS_ERRORTEXT [TRUE|FALSE]                 |
|      Write a list of English translations of all possible |
|      CLOGP error levels to standard output and exit.     |
|      Format is: $I|.          |                          |
|                                                          |
| $ clogp -PCMODELS_ERRORWORST [ERROR_LEVEL]               |
|      For TDT output, suppresses any output above the     |
|      given ERROR_LEVEL threshold.  The default is '999', |
|      eg. all results are shown.                          |
|                                                          |
| $ clogp -PCMODELS_TABLEOUTPUT [TRUE|FALSE]               |
|      Set output format to verbose tables.                |
|                                                          |
| $ clogp -PCMODELS_DUMPFRAGDB [TRUE|FALSE]                |
|      Dump fragment database to standard output and exit. |
+----------------------------------------------------------+
```

```
Note that these command line options are Daylight options which must be
defined.
```

5. Using PCmodels with other Daylight Software

The fundamental connecting thread of all Daylight software is the chemical nomenclature SMILES. In a window environment, SMILES may be copied from other Daylight programs and pasted to the **xvpcmodels** window. SMILES and TDT files written by other programs can be read into the input text pane of **xvpcmodels**, or processed by batch programs **clogp** and **cmr**. Output TDT files can be loaded into Daylight databases.

6. PCmodels Management

PCmodels includes a facility for adding and modifying CLOGP fragment data. This "algorithm manager" does not include a user-friendly interface, but provides full access to CLOGP fragment constants.

6.1 Modifying or adding CLOGP fragment data

Temporarily modifying or adding fragment data is achieved by editing fragdb.del to include the new data. The new data will be available the next time **clogp** (or **xvpcmodels** or **clogptalk**) is started. To make such changes permanent, delete the file fradb(relno).bin, (e.g. fradb461.bin), then run **clogp** (or **xvpcmodels** or **clogptalk**) which will automatically recreate fradb(relno).bin from fradb(relno).dat and fradb.del. Finally, create an empty fradb.del fro subsequent changes.

Methods for determining the values to be added is beyond the scope of this help text. However, briefly speaking, the best way to obtain a fragment value is to compare two high-confidence LOGP values for compounds differing only in that one possesses the given fragment.

Once the values are determined, it is fairly straightforward to add them to the data files. The first step should be to assure that the fragment is not already defined. Note that if a fragment is reported missing in a CLOGP calculation, this is only relevant to a particular atom environment (e.g. two aliphatic-atom and one aromatic-atom connection). Two ways of searching for an existing fragment are (1) by inspecting the data files (tricky) and (2) running CLOGP for test compounds to try each case.

If the fragment does not exist in the data files, a name must be assigned in an "ENGLISH" line. To add a fragment value, an "FVALUE" line must be added. Also, optionally, the following parameters may be defined for the fragment: BRANCH, PROXTYPE, OHCONT, NEVERAGE, ZWITTER, OCLASS, SIGMA, and RHO. For descriptions of these parameters, refer to the CLOGP Reference Manual.

It is also possible, but not recommended, to modify "CLOGP constants" (CONSTANTs and OVALUEs).

Contact Daylight Technical Support for more information on CLOGP customization.

7. PCmodels Programming "Toolkit": clogptalk and cmrtalk

In order to support users who wish to run CLOGP or CMR from their own programs, the current version of Daylight includes the programs "clogptalk" and "cmrtalk", which are designed to be accessed as "program objects" by a calling program. This program-object method is much faster than using the C function "system()". For details, see the source files in the $DY_ROOT/contrib/src/c/progob.