Creating chemical structures for online searching with MOLKICK

Peter Brueggeman
Head of Public Services
Scripps Institution of Oceanography Library
University of California San Diego

Chemical structure searching necessitates creating a structure to be searched. The most familiar method is to create chemical structures online using host system commands. Last year saw the introduction of STN Express which changed the chemical structure searching picture. With STN Express, the searcher creates chemical structures locally on a microcomputer before going online. However STN Express supports far more than structure searching and is an all-around search aid software for STN searching. STN Express supports textual searching, structure creation and searching, display of textual and structural search results, and downloading and printing of textual and structural results. While STN Express can be used for textual searching on any databank, it primarily targets textual and structure searchers of STN databases and offers particular appeal to those structure searching STN's Chemical Abstracts Registry file.

Now, MOLKICK arrives on the chemical structure searching scene with a different focus and more ecumenical slant. Searchers of the textual Chemical Abstracts database, other STN textual databases, or any databank's textual databases are not MOLKICK's market. Only those interested in structural searching are potential MOLKICK customers. Like STN Express, MOLKICK is used to draw chemical structures offline for subsequent online searching. However, STN Express draws structures that can only be searched on STN's structural databases. MOLKICK's structures can be searched in the structural databases of five databanks: DATA-STAR, DIALOG, ORBIT, TELESYSTEMES-QUESTEL, and STN. MOLKICK translates drawn structures into any of three structural query languages (CAS ONLINE, DARC, ROSDAL) which are used by these five databanks. CAS ONLINE and DARC are familiar names for query languages being used by STN and TELESYSTEMES-QUESTEL respectively. ROSDAL query language is used by databanks offering or planning to offer the Beilstein structure database using the Softron Substructure Search System (SSSS). With a structural database like Beilstein becoming available on several databanks, MOLKICK's multilingualism will accommodate the inherent differences between them. Strictly speaking, one does not go online and search with MOLKICK as one does with STN Express. MOLKICK is not telecommunication software whereas STN Express is. MOLKICK creates the chemical structure for subsequent searching by

a telecommunication software. MOLKICK concentrates on the creation of
chemical structures and leaves the selection of telecommunication
software to individual preference. Thus MOLKICK complements the
telecommunication software already familiar to the searcher. This
radical difference between MOLKICK and STN Express expands software
options for chemical structure searchers. With MOLKICK, only a
structure creation system has to be learned. Since MOLKICK does not
telecommunicate, a new software for telecommunication does not have to
be configured, grappled with, and learned. MOLKICK is designed to add
just one capability into the searcher's existing software environment:
chemical structure creation.

This review draws upon the author's experience with both text-based
telecommunications software (DialogLink, ProSearch, CrossTalk) and
graphics-based telecommunications software (STN Express, STN
Communicator, PCPLOT III). In order to better understand MOLKICK's
relationship to other software for structure searching, refer to the
author's earlier article that appeared in DATABASE SEARCHER (vol 4, no
9, Oct 1988, pp 15-23). Entitled "STN Express: search-aid software for
chemistry", the article discussed textual and structural searching for
chemical information and compared features of STN Express, STN
Communicator, PCPLOT III, DialogLink, and ProSearch. A searcher
considering MOLKICK is advised to consider it based on individual need
and the range of available software; the earlier article discusses this
at length. The experiences and opinions of others may vary from the
author's; for the benefit of everyone, please share both in the forum
of DATABASE SEARCHER.

MOLKICK is a memory resident or TSR (terminate-and-stay-resident)
software. After loading, MOLKICK is hidden away in the microcomputer's
RAM memory where the press of a hot key pops it up over another
software. Typically the software over which MOLKICK pops up will be
the searcher's telecommunication software. Since MOLKICK ties up a lot
of RAM, it should only be used in conjunction with a telecommunication
software which usually have low RAM needs. MOLKICK ties up a large
amount of real estate (280K of RAM). MOLKICK can be released from RAM
when a hot key is pressed. Therefore, after a structural search
session is terminated, the searcher excises MOLKICK from RAM in order
to free RAM up for other software (eg wordprocessing, database,
spreadsheet). Unfortunately MOLKICK cannot utilize expanded memory.
Expanded memory is additional RAM memory beyond the main 640K RAM
memory. While additional cost is required to acquire expanded memory,
it would be very useful for MOLKICK users. If MOLKICK used expanded
memory, then the searcher could leave MOLKICK permanently loaded and
tucked out of the way in expanded memory. MOLKICK would be ready to go
at all times and its usage would be greatly simplified. Instead the
searcher has to worry about loading and unloading MOLKICK from RAM in
order to use other software on the microcomputer. Makes one just want

to reboot and forget about the whole mess! For a product that is so
well designed, this is a glaring problem that expanded memory
utilization would solve.

MOLKICK installation is very simple and well documented; configuration
options include screen color, path and filenames, hot key redefinition
(in order to avoid hot keys used by other software), printer type,
databank options, and style of presentation of MOLKICK menus.
Searchers planning to use DialogLink in conjunction with MOLKICK need
to make certain specifications during installation of MOLKICK. Set the
pathname for MOLKICK's saved searches (which will be subsequently
uploaded by DialogLink) to be identical to the pathname for
DialogLink's data directory. Specify that the preset names for
MOLKICK's saved searches have an SRC file extension so that DialogLink
will recognize them as saved search strategies. If planning to
search STN's CAS ONLINE, change the host-specific adaptation for lines
per page to 0 since DialogLink will be searching in the non-graphics
mode.

Running MOLKICK from a hard disk in conjunction with telecommunication
software is facilitated by writing an ASCII batch file in order to
avoid redundant commands each time MOLKICK is used. A typical batch
file for using MOLKICK with a graphic (PCPLOT III, EMU-TEK, TGRAF05) or
non-graphic (DialogLink, CrossTalk, ProComm) telecommunication software
would incorporate commands to load the mouse software (eg Microsoft's
MOUSE.COM), load MOLKICK's graphics device driver (eg MKDHGC), load
MOLKICK itself, and then load the telecommunication software's EXE or
COM file. When the search session is over and before exiting the
telecommunication software, MOLKICK is excised from RAM with its hot
key. When the searcher exits the telecommunication software, the batch
file then excises the mouse software from RAM. With this sequence, the
main TSR rule is followed: last in, first out. Software unloads from
RAM in the same sequence in which it entered and RAM is left clear.
The following batch file serves as an example:

```
echo off
mouse
mkdhgc
molkick
telecomm.exe (eg pcplot, dl, xtalk)
mouse off
```

The batch file above will proceed in sequence and stop after the
telecommunication software is loaded. The telecommunication software
should not be instructed to go online by the batch file or software
scripts since structures would usually be created offline. MOLKICK can
be used either online or offline but it makes economic sense to use it
offline. After the batch file loads the telecommunication software

(which will then be waiting to go online), the searcher pops up MOLKICK
with its hot key combination.

MOLKICK overlays the telecommunication software with its workscreen
(figure 1). MOLKICK's workscreen has menu names listed on the top and
bottom of the screen. The menus are accessed by moving the mouse to
the menu name which causes the appropriate menu to pulldown or dropdown
(depending on installation option). Figure 1 shows the "MOVE" menu
which is used shift or rotate entire structures or fragments of
structures drawn on the screen. For experienced users, common commands
in several of the menus can be directly invoked with speed keys. Top
menus are used to specify bond types, move or rotate structures, resize
all or part of structures, copy (duplicate) fragments for building
structures block-by-block, save structures, alter the screen display
(enlarge, reduce, center) or print the screen display, change the
display of atoms (numbering, elemental symbols), and allow names or
comments to be input.

MOLKICK offers useful function key assignments. F1 is a HELP key
accessing several concise pages of ready-reference facts. F2 reveals
overlapping atoms in a structure which can then be merged. F3 turns
the display of the next higher level on or off when working with nested
Markush structures. F4 leads back to a mother structure when working
with nested Markush structures. F5 undoes the most recent structural
change; it is a UNDO key which allows one to back up one step! This is
a supremely helpful feature lacking in STN Express; nobody is perfect.
F6 erases the structure displayed onscreen; F5 will unerase it (again,
very considerate). F7 converts the displayed structure into a ROSDAL
search string for subsequent uploading to a databank using the Softron
Substructure Search System (SSSS). F9 converts the displayed structure
into a CAS ONLINE search string for subsequent uploading to STN's
structural databases. F10 converts the displayed structure into a DARC
search string for subsequent uploading to TELESYSTEMES-QUESTEL. SHIFT-
F7 removes all free sites from all atoms in the displayed structure.
SHIFT-F9 provides all atoms in the displayed structure with the maximum
number of free sites. ALT-F3 through ALT-F10 creates non-aromatic
rings of sizes 3 - 10. F8 is unused.

Attributes associated with each atom in a structure are selected from
an atom menu that pops open when an atom is created (figure 2). The
atom menu displays the atom number and the searcher can then change
atom attributes from a default setting. The elemental symbol can be
changed from the carbon default. In order to simplify structure
creation, MOLKICK contributes fifty shortcut symbols (eg MEthyl,
PHenyl, OH/hydroxy) that can be used as elemental symbols; STN Express
offers 50 also. Markush generic groups (eg alkyl, cyclic, aryl) are
also available as elemental symbols for usage on those databanks using
Softron Substructure Search System (SSSS). MOLKICK also supports atom

variables (eg all halogens, any atom except C or H) like STN Express
does. A listing of these shortcut and generic symbols is available
onscreen by pressing the F1 (HELP) function key. The searcher can use
the atom menu to lock in a specific element for subsequent structure
building. From the atom menu, other attributes can be specified like
charge, free sites, radical, mass, and valency.

The menu names appearing at the bottom of MOLKICK's workscreen refer to
saved files of chemical structures. Saved structures can be used as a
starting point for structure creation or merged into a structure being
built. MOLKICK comes with some predefined structural templates (rings
and amino acids) and the searcher can add many more. MOLKICK supports
up to four saved files in use at a given time; however an unlimited
number of files can be saved. Up to 799 structural templates can be
stored in one file.

Structures are created quite easily with MOLKICK as they are with STN
Express. Both software have mechanisms both in common and dissimilar;
creating structures with one software can be learned as readily as with
the other. Personal preference dictates which software one would
prefer for creating structures. There are differences in style and few
absolutes (MOLKICK's UNDO feature being one). After a structure is
created, it is saved as a text file (search string) using the
appropriate function key corresponding to the databank being searched.
For example, when the F9 function key is pressed to save the structure
illustrated in figure 2 for a subsequent CAS ONLINE search, MOLKICK
produces a file containing the following search string:

```
    set pagelength scroll
    STR
    .
    SET BOND N
    GRA C9,9 4,5 1,3 C8,16 1,13 2
    NOD 14 O
    BON 1-2 1-16 1-5 4-3 13-14 16-15 16-17 15-14 SE
    CON 17 E1
    CON 9 8 7 6 12 11 10 15 14 E2
    CON 1 4 5 3 2 13 16 E3
    END
```

The most elegant usage of MOLKICK would be in conjunction with graphic
telecommunication software like PCPLOT III, EMU-TEK, and STN
COMMUNICATOR. These software are capable of displaying the search
results (retrieved chemical structures) with precise structural
representation rather than with the limited textual representation seen
with non-graphic telecommunication software (DialogLink, CrossTalk,
ProComm). However since structural search results and corresponding
references are displayed much faster in textual mode than graphic mode,

there is economic merit in foregoing a graphic display of search
results. Once online, MOLKICK's saved file can be uploaded by the
telecommunication software. MOLKICK can also send the search string
without file uploading through a keyboard buffer invoked by a hot key.
The above sequence of CAS ONLINE commands will quickly build a
structure online using STRUCTURE's expert mode. The first line "set
pagelength scroll" is necessary for searchers using graphic
telecommunication software (PCPLOT III, EMU-TEK, STN COMMUNICATOR)
which have a paged screen display (no continuous scroll). The file
above indicates that MOLKICK builds CAS ONLINE ring structures using
linked chains rather than rings. CAS ONLINE's own structural display
will not look as elegant as MOLKICK's structural display. Therefore,
subsequent online modifications to a MOLKICK structure are more readily
accomplished with MOLKICK rather than with CAS ONLINE commands.
Searchers can more easily work with the MOLKICK display of distinct
rings rather than the CAS ONLINE display of linked chains.

For CAS ONLINE searchers planning to use DialogLink in conjunction with
MOLKICK, DialogLink can only recognize one databank prompt (eg Dialog's
question mark or BRS' colon). CAS ONLINE structure searchers are
presented with STN's arrow (=>) prompt and a colon prompt which appears
during CAS ONLINE's expert structure mode. In order for a MOLKICK-
created file to be uploaded with DialogLink's typeahead buffer,
DialogLink has to be configured so that the colon prompt of the expert
structure mode will be recognized. However this configuration will
cause DialogLink to ignore CAS ONLINE's arrow prompt which appears
after logon and before the structure mode is entered. Therefore, a
DialogLink autologon macro has to be created to carry the searcher past
logon and the STN arrow prompt and into the structure mode and its
colon prompt. The example following is for access via STN's Telenet
address with "#" representing STN port and "*****" representing
passwords:

```
          @\D4\C\D4D1\C\D8C 61421#\C\D8\1
          \D4********\C\2
          \D8********\C\3
          \D8\C\D63\C\D8\D8\4
          \D8FILE REG\C\D8\5
          \D2STR\C
```

With this autologon macro script, DialogLink will carry the searcher
straight into the expert structure mode of CAS ONLINE (the script can
use some fine tuning in the delay timing). The saved MOLKICK file is
loaded into DialogLink's typeahead buffer using the F7 (DISK) function
key. The first two lines of the saved MOLKICK file ("set pagelength
scroll" and "STR") can be deleted (press CTRL-Y twice) since they
duplicate commands in the DialogLink autologon macros ("3" and "STR").
The STN arrow prompt will reappear after the last line (END) of the

MOLKICK file is uploaded. Then, DialogLink will not recognize the
arrow prompt since only a colon prompt is recognized. Therefore
DialogLink's typeahead feature has to be turned off (via the F9 key)
after the "END" command has processed. An quicker process for MOLKICK
file uploading is to use MOLKICK's hot key to send the search string
through the keyboard buffer. In structure mode, CAS ONLINE rejects the
first two lines of the saved file as invalid and they cause no harm.

MOLKICK and STN Express are quite different products in intent and one
does not diminish the other. Both can be used by inexperienced and
experienced searchers for chemical structure creation. STN Express is
well-suited as a standalone package for the textual and structural
searcher who is primarily interested in searching STN's databases. STN
Express can guide inexperienced searchers in the creation of textual
searchers and MOLKICK cannot. STN Express offers advanced textual
search capabilities for experienced searchers and MOLKICK does not.
STN Express telecommunicates and MOLKICK does not. STN Express users
do not have to purchase telecommunication software. MOLKICK is best
suited as an adjunct structural software complementing the searcher's
existing software. Searchers may be already quite happy with their
telecommunication software and do not want a second one. MOLKICK does
not force a telecommunication software upon the searcher as STN Express
does. Structural searchers using non-graphic telecommunication
software will be particularly interested in MOLKICK. Some searchers
may prefer the speed and extended capability present in non-graphic
telecommunication software. Search aid software like DialogLink and
ProSearch are shining examples. Searchers can add on MOLKICK's
structural creation ability and have the best of both the graphic and
non-graphic worlds of structure searching. MOLKICK will search several
databanks' structural databases; STN Express can only search STN's
structural databases. Hard disk space may be a concern for some;
MOLKICK consumes about half a megabyte and STN Express consumes almost
2 megabytes.

MOLKICK was written by Softron GmbH, a German computer software
company, in cooperation with the Beilstein Institut. For searchers
accessing STN's structural databases, software updating of STN Express
versus MOLKICK is a concern.   When the day comes that the STN databank
makes a major change affecting structure searching software, one
expects STN to quickly revise STN Express assuming that it is still
being marketed. A MOLKICK purchaser must rely on MOLKICK's continued
presence in the market and the joint commitment of Softron and
Beilstein to update it as quickly as possible when necessary.
Interestingly, STN COMMUNICATOR offers solid performance at low cost
yet no longer appears in STN marketing literature; it is already
becoming an orphan.

MOLKICK runs on IBM PC/XT/AT/PS2/386 or 100% compatibles; a turbo PC or

an AT class microcomputer is recommended. PC/MS DOS 2.11 or later is
required. At least 512K of RAM is required. If a telecommunication
software is inordinately memory-hungry, then 640K is required. Molkick
will run from one floppy drive in all the usual 5.25 inch and 3.5 inch
sizes. A hard disk drive is not required but is recommended.
Monochrome or color monitors are supported with these graphic cards:
IBM EGA, IBM CGA, IBM VGA, IBM MCGA, Olivetti/AT&T-6300, Compaq
Portable, Hercules, Toshiba Portable. Standard printers supported are
IBM graphics needle printer, Epson 8-dot matrix printer, NEC 8 or 24
needle printer, HP Laserjet, Canon Laser Printer (LBP); unlisted
printers can usually emulate one of these. A Microsoft or equivalent
mouse is required as well as a modem and telecommunications software.
MOLKICK costs $565; For comparison, STN Express costs $595 with an
academic rate of $476.

MOLKICK is available from Springer-Verlag New York, Electronic
Information Services, 175 Fifth Avenue, New York, NY 10010 at telephone
(212)460-1622 (ISBN: 3-540-14056-5). A free working demonstration disk
that cannot create search strings for online searching is available.
It is worth a look if you need convincing of MOLKICK's utility. Check
it out!

STN Express is available from STN at their well-known address. PCPLOT
III is available from MICROPLOT, 659-H Park Meadow Road, Westerville,
Ohio 43081. Phone number is (614)882-4786. EMU-TEK is available from
GRAPHIC INNOVATIONS, 10801 Dale Street, Suite M-2, PO Box 615, Stanton,
California 90680. Phone number is (714)995-3900. TGRAF05 is available
from Grafpoint, 4300 Stevens Creek Blvd, San Jose, CA 95129. Phone
number is (408)249-7951.