CLOGP USER GUIDE

Section 1: Tutorials

Introduction to ClogP

Since its introduction in 1982, the ClogP program for calculating log $P_{oct/water}$ from structure has become the standard in the fields of pharmaceutical and pesticide design and in environmental hazard assessment. It is based on a simple but rigorous programmable-method of dissecting a solute structure into chemically-meaningful fragments. The program then accesses a table of fragment values measured in any one of five bonding environments which accounts for differences arising from connecting bond type (aliphatic, benzyl, vinyl, styryl or aromatic). If the fragment is found, but not in the desired bond environment, it estimates the difference and reports it as 'derived'. It then sums fragment values and makes important interaction corrections when fragments occur in proximity to one another. This proximity may lead to internal hydrogen bonding (+), or to the reduction of electron density of a H-bond acceptor (+), or to the steric twisting of a polar group out of conjugation (-). If the desired fragment structure is absent from the database, the program calculates it using a method (FRAGCALC) which examines all atoms and their neighbors and uses an equation based on eleven possible parameters which can raise or lower their hydrophobic contribution. A suggested reading list follows in the References section of this guide. See reference 1 for the theoretical background of solvation parameters, reference 2 for details of the ClogP algorithm, and reference 3 for some examples of practical applications.

Getting Started

The Macintosh and Windows versions of ClogP are driven from the same menus, and so the basic instructions apply to both. Manipulation of the windows and commands via keystrokes may differ, but the user will be familiar with them from prior experience.

Entry of Compounds by Name Using 'Lookup'

The 'Lookup' menu offers several choices for structure entry**. Name** allows entry of over 10,000 commonly encountered chemicals. Enter the names of four compounds in the input window, as shown below. Note that the case of the names you type is not important:

To calculate the logP for AZT, select "Terse" from the 'Options/Settings' menu, then click on the first line and select "Run one" from the Run menu. Calculation results appear in the output window. The calculated value is listed first, followed by measured value (for neutral solute), then the name of the compound. On the line beneath is a warning, indicating that the result has an unknown amount of uncertainty associated with it (ClogP always informs you of the degree of error in its calculations). The error (i.e. the deviation from measured logP) in ClogP is +0.01.

Now repeat the "Run one" procedure for the other three compounds, and consider the following progression of results:

morphine   deviation of +0.17

codeine   deviation of +0.16. The depiction can be made full screen by clicking the small square in the upper right corner of this window.

heroin   deviation of +0.08

To view the details of a ClogP calculation, you select the **Detailed** item from the **Options/Settings** menu before running the calculation. Then add three new names to your input window:

purine
adenine
methotrexate,diethyl ester

(spaces and punctutation do not matter in names–the last compound could have been entered "methotrexate di-ethyl ester")

Try entering these in the following manner: type all three lines, then select the three lines (click-drag with the mouse) then "Run one" as before. Calculation for purine is performed, and the entry for purine is de-selected in the input window; the next "Run one" will act on adenine, and so on. **cmd-R** is therefore a convenient way of stepping through a long list of entries.

purine
  Notice how the intrinsic hydrophilic character of aromatic nitrogen fragments is partly cancelled by electronic interaction within rings (+1.26) and between rings (+0.595).

adenine
  The intrinsic polarity of the amino group is negated when attached to an electron deficient ring. Compare its total electronic correction (+3.08) this with purine above (+1.86).

methotrexate diethyl ester
  This is a very complex calculation with a proximity effect of 1.0 log unit (between ester and amide) and electronic interactions of 4 log units. (see reference 2) The close agreement with measured must be partly fortuitous.

Entry of Compounds by SMILES

(if you are unfamiliar with SMILES notation, see the SMILES Tutorial below)

It is simple to generate the SMILES for a series of analogs by copying and pasting text. This makes it easy to examine the variations in logP when a compound is modified by various functional groups.

To begin, clear out your work areas; for both the output and input windows; choose "Clear home" (**cmd-K**) from the Edit menu. Set your output level to "Detailed" (on the Options/Settings menu).

Type the SMILES **NC(=O)c1ccccc1** followed by a **RETURN** (*important!*) This enters the parent benzamide structure; any additions to this line will be at the ortho position, enabling one to examine both the electronic and the ortho correction factors which ClogP employs.

Select the entire line (make sure to include the trailing **RETURN**), copy the selection, and then paste six times. You should end up with six identical lines of text.

Click at end of second line, and add the character "**C**"**.** Move down the list, (e.g. with the 'down' arrow key) and repeat the procedure, entering **Cl**, **O**, **N** and **N(=O)=O**. Now you have the structures for a series of congeners: the ortho methyl, chloro, hydroxy, amino and nitro analogs of benzamide. Your input window should appear as follows:

```
                                                    SMILES
NC(=O)c1ccccc1
NC(=O)c1ccccc1C
NC(=O)c1ccccc1O
NC(=O)c1ccccc1Cl
NC(=O)c1ccccc1N
NC(=O)c1ccccc1N(=O)(=O)
```

As in the previous example, select all the compounds in preparation for calculating logP. As before, choose "Run one" from the Run menu to perform each calculation. Referring to the calculation details for each:

benzamide   consists solely of the fragment sum, and the deviation is small (-0.01).

o-methyl   shows a special benzyl bond factor and also a negative ortho factor (-0.40) which can be ascribed to a steric decoupling of the amide group from resonance with the ring.

o-chloro   shows a large negative ortho factor (-0.83) with a small positive electronic factor (+0.17, the same as would apply to the meta and para analogs).

o-hydroxy   shows a larger electronic factor (+0.34) together with a positive ortho factor (+0.95) probably due to a strong intramolecular hydrogen bond (both substituents being amphiprotic).

o-amino   much like the o-OH, with a slightly higher electronic correction (+0.38) and much lower H-bond factor (+0.60).

o-nitro   has about the same electronic correction (+0.36) and its negative ortho factor
(-0.90) is much like that of chlorine. This is strong evidence that the amide NH_2 is incapable of donating to the nitro's oxygens, and the effect is one of decoupling from resonance and not hydrogen bonding.

As a final exercise, try running a similar series of compounds, but this time arrange the parent structure so that para-substituted compounds are entered. As before, clear both the input and output windows. Again, enter a parent structure, **NC(=O)c1ccc(cc1)**, select it, copy it, paste it six times, and enter substituents at the end of each line, so that your input window appears as follows:

```
                                          ≡ SMILES
NC(=O)c1ccc(cc1)
NC(=O)c1ccc(cc1)C
NC(=O)c1ccc(cc1)O
NC(=O)c1ccc(cc1)Cl
NC(=O)c1ccc(cc1)N
NC(=O)c1ccc(cc1)N(=O)(=O)
```

As before, select all compounds in preparation for ClogP calculation. However, this time you will perform calcuations all at once: choose "Run many" from the Run menu. LogP is calculated for all entries, and all entries are de-selected.

This brief instruction should provide some basic steps in the operation of the program. It is hoped that it will also encourage the user to ask: "What *contributes* to the hydrophobicity of this particular solute? If I substitute group B for group A, will the change in hydrophobicity be merely the difference in the intrinsic nature of the two groups, or will other factors (electronic, H-bonding, etc.) change also?"

For a more detailed treatment of these factors, see reference 2a below. Generating a number which parameterizes a solute's overall hydrophobicity is certainly worthwhile, but gaining some understanding of the forces which influence solvation can be much *more* valuable. ClogP attempts to do this; many others do not.

References

1.   a) Leahy, D., *J. Pharm.Sci*., **1986**, 75, 629.
          b) Leahy,D., Morris,J., Taylor,P. & Wait,A., *Quant.Struct.-Act.Relat.,* **1989**, 8, 17.
          c) ibid., *J.Chem.Soc. Perkin Trans. 2* **1992**, 705 & 723.
2.   a) Leo, A., *Chemical Reviews*, **1993**, 93, 1281.
          b) Leo, A., *J. Chem. Soc. Perkin Trans. II*, **1983**, 825.
          c) Leo, A., *The Science of the Total Environment*, **1991**, 109/110, 121.
3.   a) Abraham, D. and Leo, A., *PROTEINS, Structure, Function, and Genetics,* **1987**,
                 130.
          b) Leo, A., In "QSAR in Design of Bioactive Compounds", M. Kuchar, Ed., J. R. Prous
          Science, 1992, p. 335.

Annotated Calculations

A good way to become familiar with the workings of ClogP (to see how the "correction factors" it applies have important implications in the solvation process) is to browse through the "Drugs data" file with output level set at "Detailed". There is something to be learned from almost every structure, but the following seem worthy of special comment. Either scroll through the file (it is sorted alphabetically by name) and "Run one" when the cursor is on the appropriate lines, or just enter the compound names (any of the names in **bold** may be entered.

Note: As we are constantly working to improve the fragment database and algorithm for ClogP, the calculation results given may change slightly in future releases

acebutolol
With +1.57 in aliphatic proximity corrections and+ 0.61 in aromatic effects, the deviation of 0.01 is better than can normally be expected.

acronycine
The pyridone dipole correction of -1.0 slightly overcorrects in this case; deviation = +0.21

aspirin
One might expect an internal H-bond with the carboxylic acid group as donor, but the 'Ortho' correction is *negative* which indicates a decoupling from the ring, rather than positive as it is in salicylic acid.

atropine
The factor that ClogP uses to account for the proximity of polar fragments in acyclic situations seems to account well for the changes in *dipolarity* resulting from crowding, but it does not fully account for intra-molecular-H-bonding when it occurs. Here the interaction of OH and ester is given a YCCY correction of 0.78 log units, but it appears to need an additional +0.51

azaperone
Sum of proximity and electronic corrections = +1.54; deviation = +0.18

carbocloral
Proximity sum = >4.0; deviation = -0.07

Cefaclor
Uses zwitterion factor (-2.27); deviation = -0.35

chloramphenicol
The deviation of -0.14 is quite acceptible

cloxacillin
Electr. = +0.79; ortho (steric) = -1.21; deviation = +0.10

dideoxy-cytidine
Deviation = -0.01. Like its parent, **cytidine** (dev. = -0.05), the hydroxy goup on the sugar moiety cannot interact (long-range) with the nitrogen(s) in the **cytosine** ring. In contrast, **adenosine** needs an additional +1.0 as a 'conformational proximity' compared to its parent, **adenine** (dev.= +0.20)

diltiazem
The large deviation (-0.85) *may* be due to measurement error; that is, the pKa used to calculate the unionized value is much lower than expected (7.7). We believe the values in logP* database are the best currently available, but some of the deviation may arise from the measured rather than the calculated values. See also **isotretinoin**, where the measured log P might be higher

than 6.30 if carefully done by slow-stir procedure. This might eliminate the present deviation of -0.44.

erythromycin A
Based on topology, ClogP applies 11.0 log units of proximity correction, but still falls more than one log unit short (dev. +1.07). NMR and X-Ray studies show the cladinose and desosamine rings overlap, allowing intramol H-bonding at a great topological separation. With the same polar rings attached linearly off the macrolide ring, as in **niddamycin**, the deviation is lower and of opposite sign (-0.36).

ethrane
Proximity sum = 6.04; deviation = -0.36

ketamine

The restricted access to H-bond acceptor sites (β) is the probable cause of the -0.75 deviation Here water is thought to have easier access to the oxygen and nitrogen than does octanol, and it is the *competition* which determines the log P. Molecular mechanics programs may provide a means of accounting for this behavior.

LSD
The deviation for LSD is not too great (+0.26), but there is a question about the accuracy of the measured value.

methadone
The log P for the neutral form depends upon the pKa of this solute, and the values reported range from 8.25 to 10.12. Using the most reliable gives a log P of 3.93, and a ClogP deviation of -0.24 is reasonale.

quinine
Deviation is -0.15 from the measured value at pH 7.4 and with ion correction from pKa of 8.05.

saccharin
This fragment value was taken from di-aromatic attached solutes, and thus the deviation of +0.19 is better than one might expect.

serotonin
Has two pKas: 9.1 and 9.8, and some charge may remain at all pHs. Perhaps a small zwitterion factor is called for, since present deviation = -0.55. Note that the deviation for tryptamine, which lacks the phenolic OH, is only -0.07.

streptozocin
Calculation details for this structure show the eight topological proximities requiring corrections totalling 6.55 log units. The very small deviation of -0.03 is partly fortuitous.

Structure/Activity Searching Tutorial

Biological Activity Searching

To search for structures with desired type of activity, one should understand and use the permuted listing of activity types provided. Except for the comments enclosed in brackets, each word in the activity entry appears in the alphabetic listing. For example, anesthetic can be searched in a general fashion, or, as the permuted list shows, it can be narrowed to 'intravenous anesthetic' or 'inhalation anesthetic', and these can also be looked up under 'i......'.

> NOTE: For clarity, dashes are used in many activity types, such as 'anti-bacterial'. Dashes should be ignored (no space) when typed in the entry window; e.g. antibacterial. Of course, entering bacterial would result in the same hits.

The permuted list can also be useful as a dictionary when one encounters unfamiliar enzyme or receptor terminology. The abbreviation, 5ht, is fairly easily translated as '5-hydroxytriptamine' or 'serotonin'. The following examples are some of the more obscure abbreviations which can be translated from the permuted list, saving, perhaps, a search of a biochemistry text.

Looking up:

a#	finds	adrenoceptor (number)
acat	finds	acetyl-coa-acyl-tranferase
acche	finds	acetylcholinesterase
ace	finds	angiotensin-converting-enzyme
achr	finds	acetylcholinesterase receptor
ampa	finds	(isoxazole-amino-acid; find by name)
bz#	finds	benzodiazepine receptor (number)
cck	finds	cholecystokinin receptor
comt	finds	catechol-O-methyl-transferase
cox#	finds	cyclooxygenase inhibitor (number)
crf	finds	cortitropin release factor
csaid	finds	cytokine-suppressing anti-inflammatory (drug)
d#	finds	dopamine receptor (number)
dhp#	finds	dihydropeptidase (number)
eea	finds	exitory amino acid
fascioli	finds	fluke (e.g. liver fluke)
ftase	finds	farnesyl transferase
gaba	finds	gamma-amino butyric acid
h#	finds	histamine receptor (number)
hmg	finds	3-hydroxy-3-methyl-glutaryl
lhrh	finds	leutenizing hormone releasing factor
lt@#	finds	leukotriene (letter-number)
m#	finds	muscarinic receptor (number)
mglur	finds	metabotropic glutamine receptor
mmp	finds	matrix metalloproteinase
mao	finds	monamine-oxidase
nk#	finds	neurokinini receptor (number)
nmda	finds	N-methyl-d-aspartate
npy	finds	neuropeptide-y
pde	finds	phosphodiesterase
pdgf	finds	platelet-derived growth factor

ptk	finds	protein-tyrosine-kinase
tnf	finds	tumor necrosis factor
txra	finds	thromboxane-receptor-antagonist

With the Lookup set at Activity, entry of any of the above finds the structures associated with these biological functions.

Of course a disease as well known as Alzheimers is found under 'Alzheimers' or 'treatment of Alzheimers'. Gauchers disease or 'Giles de la Tourette syndrome' is an example of a more obscure ailment. The latter can be found by entering 'Tourette' and 'syndrome' In contrast to the 42 compounds used in treatment of Alzheimers, Haloperidol is the only treatment listed for Tourette's and only cycloserine for Gaucher's. While the permuted listing is no substitute for a medical dictionary, one can search for compounds useful in many treatments NOT under specifically named diseases; for example 'treatment of: cystic fibrosis, hyperpigmentation, multiple sclerosis, osteoporosis etc.

The user is advised to scan the permuted listing before entering an activity type in order to optimize the search with the correct 'String'. If one were interested in compounds used in the treatment of prostate problems, the string: prostate could be entered, which yields 6 hits. The permuted list shows that another related title is 'prostatic hypertrophy', and that entry yields an additional five hits. If the entry were truncated at prostat it would also find 'prostatic' present in (unwanted) 'leprostatic' and return 22 hits. Another example would be an interest in drugs treating brain disorders. The word, brain, appears in only one activity type, but cerebral appears in six types (several undoubtedly redundant) but producing 16 hits, and cerebro in one other.

Another example might be an interest in dopamine antagonists. To cover the entire field one enters: dopamine and antagonist and gets 34 hits. If one wants to restrict this to the d4 receptor, one inserts 'd4' after the 'and', getting 16 hits. Substituting d2, one gets 6 hits. Frequently mechanistic information is of value. Entering anticonvulsant yields 132 hits, but when and gaba is added, only three of these are known to be gaba re-uptake inhibitors. Covering an extensive field, like antibacterial, might result in an unwieldy number of hits (>600). Refering to the permuted list, one may see ways to restrict this, such as and veterinary (11 hits) or and topical (8 hits). A little practice will enable the user to format the search string to optimize the resultant hits.

Don't overlook the value of simpler searches. Entering oral results in 47 hits with varied activities, such as progestogen, hypoglycemic, anti-helmintic etc. A scan of these hits may give some insight into the type of structures which can resist breakdown in the stomach and be effectively absorbed. Even searching discontinued might yield interesting insights.

Examples of Activity searches:

Note: If the 'Partial match dialog' item in the Settings menu is not checked, the hitlist will immediately appear; if checked, it will record the number of hits and give option of saving or cancelling.

Example 1:

One might wish to study the structural commonality between various anti-parkinsonian drugs. After choosing '**Activity**' from the Lookup menu, one can type in **parkinson** or **antipark** and press 'enter'. (Do NOT use dashes in Input Panel.) The first would find 41 entries in the database (two of these are parkinson-*inducers*), and the second would retrieve 39. After the file is returned, select 'terse' if not already selected. Since one wants to view structures as well as

calculations, it is advisable to size the Depiction window to take up most of the right half of the screen, which will accomodate six pictures rather well. In the menu 'Settings' choose 3 rows and 2 columns. Then in the 'Run' panel, choose 'run page.'. The output panel will now list the first six of the hits, showing C(alculated log P), M(easured log P, if any); and molar refractivity, CMR (if selected), together with name, and activity type. The depiction will also show this information except for activity type. This list and any of the depictions can be printed whenever a hard copy is desired. Using the Zoom button in the upper right of the depiction screen will make the structural details more legible.

Students in an undergraduate Medicinal Chemistry class might peruse the output in this fashion. It should be quickly evident to them that in the hitlist are dopamine analogs or precursors (as they should have expected). But in addition they find anti-cholinergics similar to diphenhydramine, and serotonin-II agonists similar to LSD (e.g. pergolide). Thus it would seem that an antiparkinsonian drug might be designed to act upon any of these three neurotransmitters. It is also useful to note the range of log Ps which enables them to be effective in the CNS. (How does all this fit with the current hypothesis that modulation of the reticular formation is the key to controlled muscular activity? Does it suggest material for an undergraduate quiz?)

There are four obvious dopamine analogs or precursors: dopa, droxidopa, dopamantine, and ciladopa. (These could have been selected out by the 'SEARCH, Extract by Name' feature described in Ex. C. below) Without active transport, it is questionable how the first two of these could effectively cross the blood-brain barrier. Another amino-acid-type antiparkinsonian, carmantadine, has an adamantyl group added to provide sufficient hydrophobicity for passive diffusion (log P = +0.47). Isothazine and piroheptine are very similar to the tricyclic antipsychotics (e.g. chlorpromazine) which are presumed to act as dopamine *blockers.* This is indeed strange, because these antipsychotics, with prolonged use, can *cause* tremors similar to parkinsonism. Prodipine might also be thought to fit this general structural class. What are the side effects which must be looked for?

Several of the retrieved antiparkinsonians (e.g., bornaprine, chlorphenoxamine, and benztropine) strongly resemble the anticholinergic and antihistaminic, diphenhydramine. Also similar are biperiden, procylidine, pridinol, diphenidol and antitrem, all of which have a tertiary amine at one end and an aliphatic carbon attached to an oxygen and two lipophilic rings at the other. Is it significant that the oxygen (ether-type) in diphenhydramine is only an H-bond base, while the latter antiparkinsonians have an OH and thus are amphiprotic? Is it also significant that the tertiary amine in the antiparkinsonians is most often a part of a saturated ring? It would be surprising if these did not have the undesirable side effect of causing drowsiness, since diphenhydramine's use is now more as a sleep-aid (Sleep-Eze, Tylenol-PM) than as an antihistamine per se.

Several of the antiparkinsonians retrieved contain an adamantane ring or are azabicyclic. Three (benztropine, mazaticol, and tigloidine) bear some resemblance to the well-known anticholinergic, atropine.

The three LSD analogs, terguride, epicriptine and pergolide, are probably serotonin-II agonists also. One presumes that the dosage to alleviate parkinsonism is lower than that which might be hallucinogenic.

    Example 2:

A more extensive study can be made of the ACE (angiotensin converting enzyme) antagonists. The search is begun by typing **angiotens** in the input panel with the 'Lookup' menu box set at 'activity', and 62 hits result. The calculation and depicting steps are the same as in Example 1.

Thirty six of these structures have acquired names and have been subjected to clinical testing or have been approved as anti-hypertensives. The balance of the structures in this file have been given company research names, and their activity or their side-effects may not render them suitable as pharmaceuticals. It is interesting to note that all but seven of the named ACE inhibitors contain the feature: $HO_2C-CH_2-N-C(=O)-$; that is,. an alpha-amino acid with the amino nitrogen part of a carboxamide. In all but delapril the nitrogen atom is part of a ring. Is this feature a true 'pharmacophore'?

Some of the structures with company research names (PD-123,177, PD-121,981, MDL-27,788, MDL-27,088, FPL-63,547XX, AB-47, SQ-29,852) also contain the amidized amino acid noted above. At some point in time, ACE-inhibitor research seemed to take another direction. The amidized nitrogen was replaced with a nitrogen in imidazole, and often a second acidic group (tetrazole) was added. In this category one finds: EXP-3174, DUP-532, GR-117,289, CI-996. In S-8308 a methylene is inserted between the imidazole and the carboxyl, and in L-158,978 pyridine substitutes for the imidazole nitrogen. The distance between carboxyl and the unionized nitrogen may not be crucial, as seen in A-81,988 and Candesartan. Indeed, the carboxyl may be superfluous, as seen in RA-633, L-158,809, and L-159,093.

This study may give some insight into the rationale used by CADD departments as they sought more active and diverse ACE-inhibitors. It also may provide some interesting raw material for training students in the use of the 3-D overlay procedures commonly employed in molecular mechanics programs and CoMFA.

    <u>Example 3:</u>

In this example, we study Anti-Helmintics (Anthelmintics) first using Lookup by Activity as in Examles 1 & 2, and then by using the 'SEARCH: extract from Window' feature as a rough substructure search tool.

Using Activity under Lookup, enter **helmintic** and follow the procedure for listing and calculation given in Example 1. The search finds 103 compounds with anthelmintic activity, which is a rather large group to study in detail for structural similarities. One quickly notices that a number of the names contain 'AZOLE'. Choose 'Extract from Window' from the Search menu. On the line indicated, type in AZOLE (case sensitive here) and 18 azole-containing structures are returned. Eleven of these (lobendazole, parabendazole, albendazole and its oxide, oxibendazole, dribendazole, fenbendazole, oxfendazole, luxabendazole, mebendazole, and etibendazole) have an alkyl carbamate moiety at the 2 position of benzamidazole, and all but the first have a hydrophobic group at position 5. Is this some sort of 'pharmacophore'? In cambendazole the carbamate has been moved to the 5-position and a thiazole placed at 2, and thiabendazole does away with the carbamate altogether. Furodazole has a furan at 2- and a pyridine ring fused onto positions 6-7. To this point the study is interesting from a historical point of view, indicating perhaps that the first ten were produced to take advantage of an early lead, only to find that features that appeared at first to be essential really were not. Further study reveals an unusal diversity of structures, indicating, perhaps, that the worms (helminths) may be attacked by a number of mechanistic approaches. Certainly invermectin (choice for eye infestations), kainic acid, and nitramisole must act in different ways.

    <u>Example 4:</u>

In this example we examine string search by Name. In assigning names to promising pharmaceuticals, USAN has tried to consistently include a 'clue' as to structural type or activity. This can be useful on occasion.

(a)   Select 'Name' from the Lookup menu, then enter **prost**, which returns 72 hits. With a few exceptions (e.g. prostaphilin), all are prostaglandins. Some of the five membered rings contain carbonyls while others are hydroxy-substituted, and still others are fused to other rings. This makes it difficult to 'capture' all the prostaglandins by any substructure search method. Entering **prostagland** in an Activity Type search results in only 8 hits, since the more specific activity has been entered as more informative (e.g. abortifacient, anti-ulcer etc.)

(b)   Repeating a name search with **profen** one gets 45 hits, most of which are the phenylpropionic acid type of NSAIDS. Some of the non-targeted compounds may be thought-provoking, such as diprofene (anti-cholinergic) and iprofenin (diagnostic aid). Profenophos (insecticide) is obviously an outlier to this class.

(c)   Repeating the search with **pride**, one gets 62 hits. Under 'Tools', select **'Add as comment: activity'**, and one sees that the 'pride' suffix applies to a variety of CNS activities, such as: dopamine antagonists, anti-migranes, anti-psychotics, anti-emetics, tranquilizers, etc. This is a feature which could be useful to students of pharmacy.

Examples of Structure searches:

ClogP also features a structure/substructure searching routine of moderate capability which is based on an 'invisible' WLN (Wisswesser Line Notation).

   Example 1:

In the 'Search menu', draw down to 'Structures' and slide to the section of the alphabetic listing of substructural types. As an example, one could click on **cephalosporin** in the 'A to C' section. This returns 192 hits, sorted by name, in a new window labeled 'Matched T46 ANV' in Master'. These can be calculated and scanned six structures at a time as explained above.

   Example 2:

Using the 'Tools' menu bar it is easy to find the kinds of activity associated with a given heterocycle, such as pyridazine. Again select 'Structures' from the Search menu, and drag the mouse to the middle alphabetical group and release on **pyridazine**. This returns 145 hits with the WLN in the comment field. From the Tools menu, select '**Delete comments'**. After this is complete and **Name** selected in Lookup , select '**Add comments'**, and drag to '**activity'**. When this operation is complete, select 'Sort on comments'. The structures without activities will all be at the top of the window listing the hits, and these can be deleted. The remaining 19 show that the pyridazine ring appears in solutes with the following activities: analgesic, anti-bacterial, anti-depressant, anti-hypertensive, anti-viral, herbicide, MOA inhibitor, muscarinic agonist, and vasodilator.

   Example 3:

In Example 1 one might have wanted to find cephalosporins which also contained the tetrazole substructure. This more restricted search can be performed if one accesses the WLN Dictionary directly through 'Open File'. After copying the WLN-tag for cephalosporin and pasting it in the Input Panel, type in 'and', return to the WLN Dictionary and copy the WLN-tag for 'tetrazole, 1-subst.' and paste it in the input panel. Be sure the menu panel 'Lookup' shows 'WLN' and 'Enter' or 'run one'. This finds 39 cephalosporin analogs which also contain the tetrazole moiety. If one often needs to search using a second substructure which has a simple WLN label, one might want to jot a few of these down so that direct access to the

dictionary is not required. For example, searching for **phenothiazine** returns 125 hits, and 'SEARCH: Extract from Window' can be used to find interesting subsets. One finds 16 analogs with the CF$_3$ (**XFFF**) group; 4 with cyano (**CN**); but none with nitro (**NW**). There are 20 with piperidine rings (**T6NTJ**), 25 with piperazine (**T6N DNTJ**), but only one with morpholne (**T6N DOTJ**). There a five with acidic groups which result in zwitterions: 2 sulfonic acids (**SWQ**) and two carboxylic acids (**VQ**). The possibilities for interesting searchs seem endless.

SMILES Tutorial

Essential SMILES

Simplified **M**olecular **I**nput **L**ine **E**ntry **S**ystem has 4 basic rules which cover 98+% of all entries. These can be learned in a few minutes, but of course it will take some practice to become proficient.

Rule 1
Use ordinary atomic symbols

- B, C, N, O, P, S, F, Cl, Br, I --used as shown
- suppress hydrogen **except** on pyrrole nitrogen where it is [nH]
- Other atoms and any charges are placed in brackets; e.g. [Si]; [N+]; also any hydrogens to fill valences of non-ordinaries; e.g. dichlorosilane Cl[SiH2]Cl
- Use upper case for aliphatic; lower case for aromatic. But second letter of any symbol is always lower case.

Examples:

CH_3CH_2OH	CCO
$CH_3CH_2NHCH_3$	CCNC
pyridine	n1ccccc1 (ring numerals explained in Rule 4)
pyrrole	c1c[nH]cc1
$Pb(CH_3)_4$	[Pb](C)(C)(C)(C) or C[Pb](C)(C)C

Rule 2
Bonds need not be specified

Exceptions are:
- Equal sign (=) for double bond.
- Pound sign (#) for triple bond.
- A period separates disconnected structures; e.g. ions or complexes
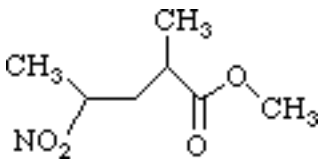
Examples:

acetonitrile	CC#N
$CH_3CH_2COO^-Na^+$	CCC(=O)[O-].[Na+]
Azido substituent	*N=[N+]=[N-]
ammonium ion	*[N+H3]
SO_2NH_2	*S(=O)(=O)N
$HC{\equiv}CNO_2$	C#CN(=O)=O

Rule 3
A branched group is placed in parentheses

- Branches can be nested if desired.
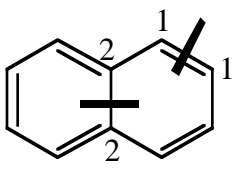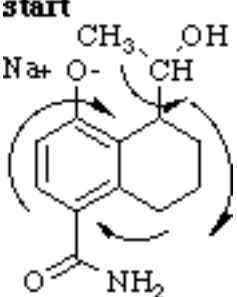- Can follow any path desired; will be made unique for storage.

Examples:

| CF$_3$CH$_2$C(=O)CH$_3$	FC(F)(F)CC(=O)C
Ph-N$^+$N·(BF$_4$$^-$)	c1ccccc1[N+]#N.[B-](F)(F)(F)F
	CC(N(=O)=O)CC(C)C(=O)OC

Rule 4

To make a ring pathway linear, one bond must be 'broken' for each ring

• One can number the atoms on each side of a break with the same number.
• Any number can be re-used after that ring is closed.

Examples:

benzene	c1ccccc1
	c1c2ccccc2ccc1
	CC(O)C1CCCc2c(C(=O)N)ccc(c12)[O-].[Na+]

Many complex compounds appear in the logP* database, and these can be entered via name. For example, **strychnine** can be typed in the SMILES input window much faster than entering the SMILES itself. However, the task is not as difficult as it first appears, if one chooses a path which first follows the periphery of the structure and then leads into the center. Beginning arbitrarily at the top of the benzene ring as shown, the lower case 'c' is followed by the number '1' showing that it will later be joined the the carbon fused to the pyrrolidine ring (that 'c' will also be followed by '1').
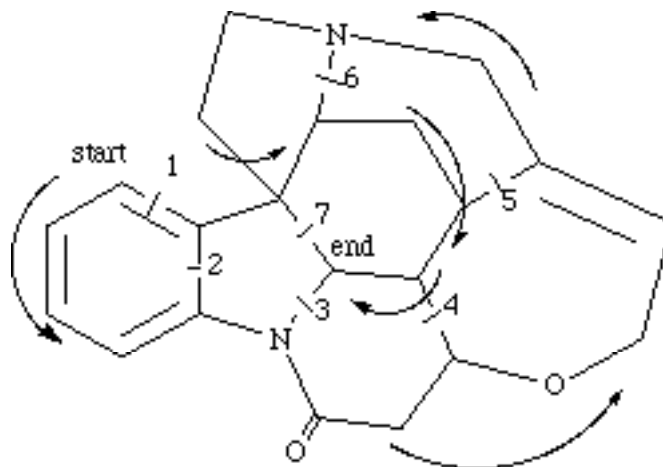
As the path follows the periphery, bonds to the interior are 'broken' and numbered in succession, until the path enters the interior at the quaternary carbon and the seventh and last 'ring break' is made. (Note that the highest number should equal the number of rings present.) The SMILES for this path is:

c1cccc2N3C(=O)CC4OCC=C5CN6CCC7(c12)C6CC5C4C37

Before beginning the calculation, it is a good practice to check to see than each numeral in the SMILES appears as a pair. Note that before the path enters the central region, it branches to make connection to the beginning carbon (at c12), Note also that these numbers refer to 'one' and 'two' and not 'twelve'. Numbering the bond breaks is sufficient to "keep score", and it is more legible in small diagrams than numbering both sides of the break as was done in the previous example. It takes a few hours of practice to become comfortable in entering complex structures via SMILES, but if one can handle the strychnine example, there are few in all of chemistry that will be more formidable.

Using ChemDraw Pro to create SMILES

CambridgeSoft's ChemDraw Pro has the ability to copy any structure as SMILES, which you can then paste into ClogP For more information on CambridgeSoft and their products, point your WWW browser to: **http://www.camsci.com/**

1. Draw a compound in ChemDraw Pro (version 3.5 or newer), labelling all non-carbon atoms. At this time, ClogP does not understand isomeric SMILES, so any included isomeric information will be ignored.
2. Select your drawing and choose "Copy As... SMILES" from the Edit menu.
3. You can now do one of the following:
    • Paste the SMILES directly into ClogP's SMILES & name input window. Continue the first 2 steps until you have the desired number of SMILES.
    or
    • Paste the SMILES into any text file. Continue the first 2 steps until you have the desired number of SMILES, then Save the text file (make sure to Save as text, if optional) and Quit the text editor. Now start the ClogP program and Open... the text file you just created.
4. Run one or Run many as you normally would.

Importing MDL MOL files

ClogP can also import and export MDL Informations Systems, Inc. MOL connection table format files, such as those created with ISIS/Draw. You can download ISIS/Draw from MDL's WWW page: **http://www.mdli.com/** or call them at (800) 635-0064 for more information.

You can import MOL files into ClogP in one of two ways:

1. Be sure that the ClogP Input window is the active window, then choose 'Import MOL' in the File menu, locate the MOL file which you wish to import and click 'Open'.

or

2. Be sure that 'Post CTfile to Clipboard' is checked in the ISIS/Draw Preferences (in the Options menu), then simply select a compound, choose 'Copy' from the Edit menu, go the ClogP application and 'Paste' into the Input window. The connection table format is automatically converted to SMILES.

Modify SMILES Tutorial

One often wants to calculate log P for a modification of a complex structure. Those familiar with SMILES will often be able to spot the portion of the SMILES string in the Input Panel which needs editing, but the following examples illustrate how the ClogP 'Modify SMILES' feature can make this easy for those less familiar with SMILES notation.

Using the Modify SMILES feature

Example 1:

As a first example, we might want to study four possible variations of the anti-neoplastic agent, **Ara-C**:

Ara-C   2'-deoxy   2',3'-dideoxy   2',3'-dehydro   3'-fluoro
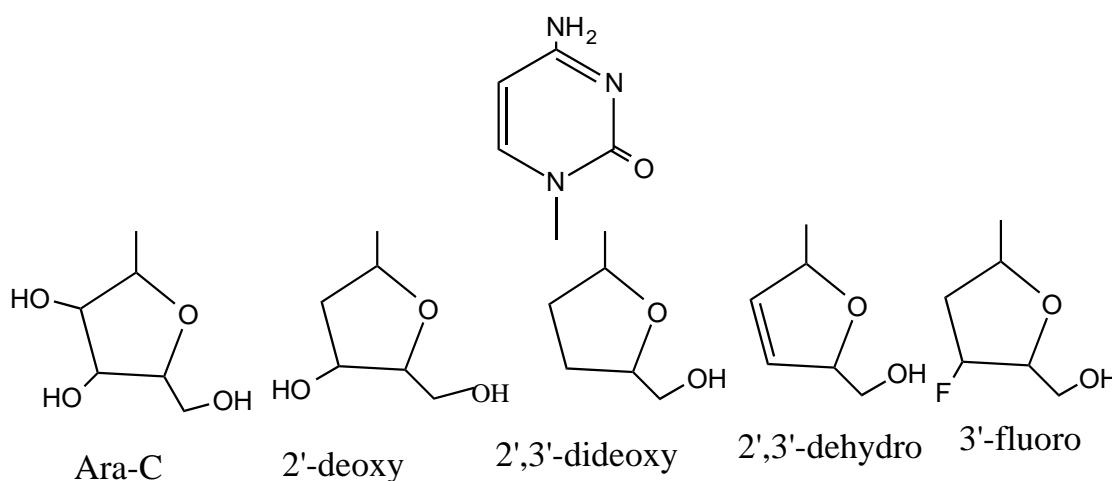
1. Select "Name" in the Lookup menu, then type the name, **Ara-C**, in the Input Panel, press 'enter' (NOT 'return') and a depiction with calculated and measured value is provided.

2. From the 'Edit' menu click on 'Paste SMILES' and the Input Panel will now replace the name with its corresponding SMILES string. Select this string.

3. Now from the 'Edit' menu click on 'Modify SMILES'. This returns a panel with the SMILES preceded by the cursor. To advance the cursor to the right, to get to the atom where the change is to occur, the Advance Key (**AK**) on the Macintosh is '**?**' and on the P.C. it is **F10**. Each time the AK is entered, the Depiction Panel highlights the atom in red which is just to the left of the cursor in the SMILES notation. In the Ara-C example, the inital point is the amino group on the cytosine ring. Upon continued entry of the AK the path follows the carbon side of the cytosine and then crosses to the arabinose ring. Note that when a parenthesis or ring numeral in the SMILES is immediatly left of the cursor on the Macintosh no atom is colored. Also note that the AK on the Macintosh ('?') automatically repeats if held down, but the F10 key on P.C. must be struck repeatedly. If the desired atom is passed by, the cursor can be moved left with the arrow key or repositioned by mouse.

4. If we wish to calculate log P for the 2'-deoxy analog (i.e. removing the OH from the carbon adjacent to the one linking the arabinose to cytosine), we enter the AK until that hydroxy is highlighted. The cursor is just to the right of the oxygen in parenthesized sequence:

..C1O|). The 'delete' key removes this hydroxyl from the depiction, and **OK** produces the new calculation.

5. If we wish to calculate log P for the 2',3'-dideoxy analog, we can again enter the 'Edit' menu and choose 'Modify SMILES'. Repeating the AK will return the cursor to the 3' hydroxyl highlighted in red in the depiction, and again the 'delete' key removes it. Using the 'enter' key shows that this analog has been named, Zalcitabine, and its measured value is in excellent agreement with calculated.

6. A further modification might be to create the 2',3'-dideoxy-2',3'-didehydro analog. In this case the cursor is stopped as soon as the 3'-carbon is highlighted and then the equal sign, =, which is the SMILES symbol for a double bond, is typed. The depiction shows the desired structure and 'enter' returns a calculation in reasonable agreement with measured.

7. A final modification might be to create the 2',3'-dideoxy-3'-fluoro analog. By this time the user may be familiar enough with the SMILES string that the editing can be done directly in the Input Panel using the mouse. Since there is only one double bond '=' to be deleted, the cursor is placed just to the right of it to delete. If the Modify SMILES window is used, editing can be either with the mouse or by AK to the point where the double bond is highlighted. After deletion, the next use of AK places the cursor within the empty () which originally held the 3'-hydroxyl, and one can now insert the flourine symbol, F. This returns the correct depiction, and 'Enter' gives the calculation which deviates -0.37 from measured.

Example 2:

A second example shows how the steroid, **progesterone**, can be modified stepwise to produce the following structural analogs: (1) the 17,21 di-hydroxy analog; (2) the 1,2-dehydro-11-keto-17,21-dihydroxy analog; and (3) 9-fluoroprednisolone.
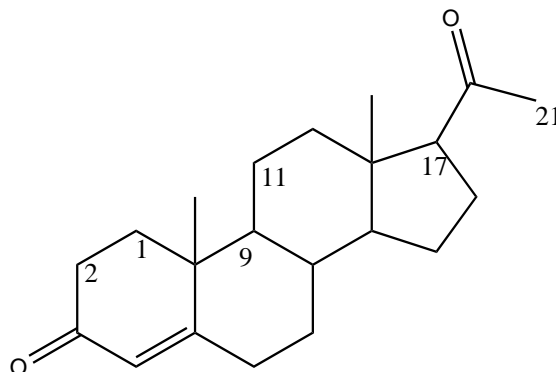
1. In the Input Panel type **progesterone**, then 'Run One', whence the calculated logP, the measured log P and the depiction are returned. From Edit, choose 'Paste SMILES' followed by 'Modify SMILES'. The first use of AK highlights the 21-carbon which happens to begin the string.. Since it has moved the cursor to the *right* of carbon 21, one needs to add the hydroxyl as a 'branch'; i.e. in parentheses as **(O)**. (Remember, it is not necessary to specify hydrogens although (OH) will work; see SMILES tutorial). The depiction confirms this step, and so continuing with AK one sees the carbon-17 highlighted when the cursor is between a C and the number '3' which indicates a ring closure. One should **never separate** an atom from its ring closure number, and so the correct position for inserting the OH has the cursor to the right of the 3, even though, on the Macintosh, the

carbon is no longer highlighted. Type in **(O),** enter (as OK), and the calculation for the 17,21- dihydroxy analog, Cortexolone, is returned. It is in good agreement with measured.

2. From 'Edit' again select Modify SMILES. Repeat AK until carbon 2 is highlighted. Insert = and continue with AK until carbon 11 is highlighted. To convert this carbon to a carbonyl, type in **(=O)** which denotes a branched double bonded oxygen. After the depiction is deemed to be correct, click on OK, which returns the calculated value for prednisone. (The calculated values for steroids require a special positive correction for polar groups (almost always OH or keto) on carbon 11. It has been postulated that octanol is unusually effective in accomodating this feature because of the 'flat' expanse of alkane area surrounding the acceptor site which the octyl chain finds 'comfortable'

3. Changing the 11-keto group of prednisone to a hydroxyl and adding a 9-fluoro group will convert it to 9-A-fluoroprednisolone. Choosing Modify SMILES, one uses AK until the 9-carbon is highlighted and then spaces past the ring closing numeral to insert **(F)**; then proceeds until the cursor is to the right of the = in the next (=O). Deletion of the double bond produces a hydroxyl symbol, and OK returns the calculation.

Example 3:

The next two examples deal with changes to a pyrrole-type nitrogen, which is the only instance where a hydrogen needs to be specified in SMILES (see Tutorial). In the first example deletions will be made on the N-ethanol chain of metronidazole, and the final example will be an addition to the ring nitrogen in serotonin.

1. Type **metronidazole** in the Input Panel, enter, and from the Edit box choose 'paste SMILES' followed by 'Modify SMILES'.

2. Entering the AK begins with the methyl substituent and follows around past the nitro group to the N1 of the imidazole. The balance of the SMILES (CCO) specifies the ethanol substituent. Since deletions are made to the left of the cursor, it is best to use AK until the hydroxyl is highlighted. If the delete key is pressed twice, the N-methyl analog will be depicted and calculated using OK. If the entire substituent, CCO, is to be removed the cursor will end up just to the right of the 'n' in the SMILES. The depiction will show this to be a pyridine-type nitrogen (as the N3) and thus invalid. One must type in 'H' for the program to use the correct pyrrole-type fragment (See Tutorial) and calculate log P correctly.

Example 4:

1. Type **serotonin** in the Input Panel, and after the paste and modify steps, use AK until the ring nitrogen is highlighted. The cursor will be immediately right of the bracket in [|nH]. A substituent on this aromatic nitrogen will eliminate the need for the brackets and the hydrogen specification, and so one steps the cursor past the right-hand bracket and presses 'delete' until both brackets are removed. (The odd depiction at this step is of no concern.) Then re-type the 'n' follwed by the substituent in parentheses: n(C) for N-methyl.

Example 5:

The previous examples dealt with alterations of a given ring structure. Adding a ring to a parent structure via MODSMI is slightly more involved. After the 'paste SMILES' and 'modify SMILES' steps are taken with the parent as in the previous examples, start the AK sequence to see the path which SMILES takes through the structure. With the mouse, place the arrow beside the first atom which will begin the new fused ring just to mark the place (don't click).

When the highlight comes to this atom, enter a ring-closure number *one higher* than the highest in the parent SMILES. If the first fusion atom already has a closure number, put the new number immediately following it. Type in the new sequence of atoms in the fused ring. Then use the right arrow to jump past the next atom (which will become the other new fusion atom) and give it the new fusion number.

1. **1,2-benzacridine** from **acridine**:
   Enter 'acridine' and edit using 'paste SMILES' and 'Modify 'SMILES'. Place the arrow beside the desired carbon in the Depiction. Run the highlight to this carbon, type in '4' followed by four lower case c's. Skip the following 'c' using the right arrow key, and type in '4'. Depict shows the correct structure ready for calculation by clicking on OK.

2. **endrin** from **aldrin** (Epoxidation of double bond):
   In 'Modify SMILES' space past the double bond in CH=CH and delete it. The epoxide will create a fifth ring, and so type '5', followed by 'O'. Then skip past the next carbon and close the epoxide ring with '5'. **OK** will return the calculation and name, endrin.

3. An acetonide from cis-hydroxyls on carbons 16 and 17 of a steroid, for example:
   triamcinolone-acetonide from **triamcinolone:** Start the Modify SMILES using AK until the carbon-16 is reached. Insert the next higher ring closure number, 5. The next oxygen no longer will be a branched OH, and so remove its parentheses. Next add an aliphatic carbon with two branched methyls; i.e. C(C)(C). Complete the acetonide ring with 'O', then skip past the next carbon and give it the ring closure, 5 as well as its original 4. The branched OH still remains, and so skip past the (O) and delete it. Entering this structure returns the name and calculation for triamcinolone-acetonide.

   When linked to Cambridge Soft's ChemDraw, an alternate (and perhaps simpler) method is to enter triamcinolone by name, copy its pasted SMILES to ChemDraw, connect the two hydroxyls at C16 & C17 with -C(Me)(Me)-, copy as SMILES, and paste back into CLOGP.

Section 2: Reference

ClogP Menus At a Glance

File	
This menu contains the standard items for file handling, and also tools for importing and exporting MDL .mol files.	
New	Open a new, empty input window (up to seven allowed)
Open...	Open an existing file. ClogP can read any 'TEXT' file
Close	Put away active document
Save	Save active document using current name
Save As...	Save active document with a new name
Save a Copy As...	Save a copy of active document under a new name
Revert	Discard changes to current document and open last saved version
Import MOL...	Open an MDL 'MOL' format connection-table file. The contents of the file are translated to SMILES.
Export MOL...	Save active document in an MDL 'MOL' file.
Page Setup	Set printing options
Print	Print current window
Quit	Exit ClogP application
Edit	
This menu contains the standard items, and also a few ClogP specific editing tools.	
Undo	Not implemented
Cut	Remove selected text (retrieve from the clipboard with 'Paste').
Copy	Copy selected text to the clipboard (or copies images if depiction window is frontmost).
Paste	Paste contents of clipboard at insertion point.
Clear	Remove selected text. Note: text is not placed in clipboard, so cannot be retrieved later.
Clear home	Remove all text from 'Compound input' & 'Models output' windows, and deletes 'current compound' (no depiction).
Select All	Select entire contents of document.
Paste SMILES	Paste SMILES string of current compound.
Modify SMILES	Interactively modify a compound (via SMILES string). See the 'Modify SMILES Tutorial' for more information.
Edit License	Modify ClogP program license. Use with caution–you must know the proper license key!
Search	
This menu provides additional search options (input windows always allow lookup by exact or partial identifier). The 'find' items let you look for text in the active program window. The 'scan' items filter text from some source into a new program window.	
Find in window	Search current compound entry window for text string.
Find Next	Repeat last 'Find'.
Extract from window	Lines containing specified text copied from current input window into a new window.
Extract from file	Lines containing specified text copied from a selected file into a new window.
Extract by name	Lines containing specified text copied from MASTER database into a new window. Tip: Use this when an exact compound name match prevents lookup of a group of similarly named compounds, e.g. 'mitomycin'.
Extract by structure	Series of menus leading to items defining structural groups. Selecting an item performs a structural search; results are placed in a new window.

Settings

This menu lets you customize the behavior of ClogP. Items set here are restored each time you run the application.

ClogP	Calculate logP model. Can be toggled on/off.
CMR	Calculate MR model. Can be toggled on/off.
Terse	Set calculation output to the minimum: a one-line summary.
Detailed	Sets calculation output to intermediate level.
Verbose	Sets calculation output to maximum possible.
Depict rows	Set rows for 'Run page' depiction, selecting from pop-up menu.
Depict columns	Set columns for 'Run page' depiction, selecting from pop-up menu.
9 point	Display text for active window in a smaller font (default).
12 point	Display text for active window in a larger font.
X-ref list	If checked, when ClogP finds no exact match for a name (or activity, etc.), it shows a dialog containing the partial matches. At that point you may select some, one, all, or none of the matches. If this option is unchecked, all matches are put in a new window.
Temporary files	If checked, ClogP makes local copies of certain search files that reside on a network server, to provide much faster 'partial match' searches. Can be toggled on/off. Note: this option is enabled only when the program is launched via the network.

Run

Calculate logP and/or MR, and generate depictions, for one or more compounds using this menu.

Run one	Calculate models for a single compound. Operates on either the current line (where cursor sits) if there is no selection, or the first line in selection, if multiple lines are selected.
Run many	Calculate models for all selected lines. Note: this item is enabled only when one or more lines selected.
Run page	Calculate models for a group of R x C compounds, where 'R' & 'C' are the 'Depict rows' and 'Depict columns' items from the Settings menu.
Run batch	Calculate models for all compounds in selected file, with output directed to a second file. Tip: Use this to circumvent the 32,000 character size limit on ClogP document windows.

Tools

This menu contains a variety of tools for manipulating the text of document windows. First, there is a series of specialized sorting tools, then options for adding/removing database information for each compound. In contrast with the Edit menu item, which operate on single lines or selected text, these items affect the entire contents of the document.

Sort	Sort based on entire line.
Sort by name	Sort based on compound name. Note: works only for names as explicitly given, not names as they appear in the database.
Sort by comment	Sort based on comment (starting with first, if there are multiple).
Sort by structure	Sort based on WLN. An alphabetic sort is done on WLN strings, which generally has the effect of grouping similar compounds.
Sort by ClogP	Sort based on calculated logP (works only for 'Model output' window).
Add SMILES	Insert SMILES before name (if not already present).
Add name	Insert name after the SMILES (if not already present).
Add as comment	Add identifier as comment at end of line. Identifiers are selected from a pop-up menu (name, CAS, etc.).
Remove SMILES	Discard SMILES (requires that a name be present).
Remove name	Discard name (requires that a SMILES be present).
Remove comment	Discard all comments (if any present).

Window

Selecting an item from this menu makes the corresponding window active. The first three items are permanent. Each item has a corresponding keyboard shortcut. The first new input window you open has the shortcut cmd-4, and so on. For further information, see the 'Windows' section below.

SMILES input	Make input window the active window.

Model output	Make output window the active window.
Depiction	Make depiction window the active window.
Lookup	
Use this menu to control the default 'lookup' type, the kind of identifier, in addition to SMILES, the program will attempt to match in the database. For example, if 'CAS' is set as the lookup type, each line you 'run' will be interpreted as a CAS number (if it is not a SMILES). The current lookup setting is reflected in the title of the default input window.	
Name	Compound name.
Activity	Activity type (biological action).
MF	Molecular formula.
CAS	CAS registry number.
WLN	Wiswesser line notation (describes compound structure).

Windows

The program starts up with three open windows, one for compound entry and two for displaying output. Each is described in detail below.

Default input window

Entry is by one of six compound 'identifiers', either a SMILES or one of: name, CAS number, MF, WLN, or activity.   The current default identifier is set via the Lookup menu, and is reflected in the window title, e.g. if CAS is the default, the title reads 'SMILES & CAS input'. See the 'Input' section below for a more detailed discussion of compound entry.

This window is considered to be a scratch 'worksheet', so you are not prompted to save changes when you quit the application. This is the exception, not the rule, for input windows–in all other cases you will be reminded about unsaved changes. (cf. 'Other input windows' topic below)

Model output window

There are three output settings: terse, detailed, verbose. They apply to both the logP and MR model output; The descriptions below apply when both models are 'on'.

The 'Terse' setting reduces calculation output to the minimum possible, a one-line summary containing:

>   calculated logP, preceded by a 'C:'
>   measured logP, preceded by an 'M:'
>   calculated MR, preceded by 'CMR:'
>   compound name (SMILES displayed if no name available)
>   error warning as comment (if any)

The 'Detailed' setting displays a map of the contributions to the final calculated value:

>   summary line +
>   contributions table

The 'Verbose' setting additionally shows an analysis of the compound's structure:

>   summary line +
>   compound name
>   fragment map (ClogP only)
>   contributions table

You cannot enter any text in this window, though you may copy out of it, and delete text from it with 'Cut' and 'Clear'.

All ClogP program windows are limited to 32,000 characters of text. While this may not seem too stringent, it does not take that many 'detailed' results to approach this limit. When this happens, the oldest material is (at the 'top' of the window) is silently discarded, to make room for the new results.

You are not prompted to save changes in this window when you quit the application. If you wish to save its contents, either copy & paste to another application, or use the 'Save a Copy As' menu item.

Depiction window

This window displays a 2-D representation of the current compound(s). Above each compound a name is given, below ClogP, measured logP, and CMR are given, as available.

On color monitors set to display 8 or more colors:

> blue   isolating carbons
> yellow polar fragments
> red   missing fragments

On greyscale displays, these three types of atoms are also drawn in a distinct shades . On monochrome (black & white) screens no visual distinction can be made–all atoms are drawn in white.

The depiction window can be put away if you do not want to see the 2-D image. To make a 'hidden' depiction window visible again, select the 3rd item from the 'Window' menu (or enter cmd-3).

Other input windows

You may open a new input window at any time; the program may also need to make one, as output from another operation. In all, there are four ways to create a new input window:

1. With the 'New' menu item. These windows are initially blank.

2. With the 'Open' menu item. ClogP only lets you open files of the proper type–if a file doesn't appear in the 'Open' dialog box, it is not a 'TEXT' file. Tip: Files can also be opened by dragging them onto the ClogP or PC icon, or with 'launcher' utilities; files previously saved by ClogP can be opened by double-clicking their icons.

3. As a result of an 'ambiguous' identifier. For example, entering 'salt' would bring up a new window listing all the compounds with 'salt' in their names.

4. As a result of a 'Extract' operation. Continuing the example above, using the 'Extract by Name' menu item with the target 'ammonium' would bring up a new window containing only ammonium salts.

These input windows behave just like the default input window, except that they can be closed (and you are prompted to save changes).

Since a total of 9 program windows may be open at any time, you have a limit of 6 user windows (7 if the depiction window is put away).

Input methods for ClogP

Identifiers

ClogP understands six types of 'identifiers', or references to compounds:

> SMILES
> name
> CAS number
> molecular formula
> WLN
> activity

The program interprets SMILES at all times, and accepts exactly one other identifier at any given time. The current default identifier is set via the Lookup menu, and is reflected in the window title, e.g. if CAS is the default, the title reads 'SMILES & CAS input'.

Note: all interpretation is triggered by commands from the Run menu.

Some identifiers are normally exact (SMILES, CAS) while others are sometimes or usually ambiguous (name, MF, activity). ClogP's response is always the same: when an exact match is found, model results are output, and depictions drawn. When the input is ambiguous, no exact match is found, a new window is opened that lists the partial matches (more precisely, exact sub-string matches). If no match is found at all, an alert dialog informs you of this.

Interpretation

ClogP 'parses' each line of input in three steps. This process may seem confusing at first, but the design allows maximum flexibility in user input. Each step is illustrated below with sample input.

> 1st try: Interpret leading word (text before the first blank space) as a SMILES. If successful, the remaining text is taken to be the 'name' of the compound. Any text after a '#' symbol is treated as a 'comment'.

```
COc1ccn(C)c(=O)c1C#N                      RICININE          # CASTOR BEAN POISON
SMILES                                    Name              Comment
```

> 2nd try: If a comment is present, interpret text prior to comment as a name (even if default identifier is CAS, MF, etc.):

```
RICININE        # CASTOR BEAN POISON
Name            Comment
```

> 3rd try: Interpret text (prior to any comment) according default identifier:

```
    POISON
    Activity identifier
```

Standardization

It is not necessary to learn any special rules for identifier input because the program standardizes your input. Rules are applied appropriate for each type of identifier. In all cases,

tabs are converted to blanks, and extra blanks (leading, multiple) removed. Also, except for SMILES, case of input does not matter–input is always converted to uppercase. Examples of raw vs. standardized input given below.

SMILES – isomeric specifications ignored (case is significant):
Br@C=CBr   —>   BrC=CBr
Br\C=CBr   —>   BrC=CBr

Name & CAS – all punctuation & spaces ignored:
2,3 dinitro benzoic acid   —>   23DINITROBENZOICACID
15147-64-5   —>   15147645

WLN & Activity – no extra standardization done!

Searching

As explained above, much of ClogP's 'search' capability is invoked automatically, with no special effort the part of the user: sub-string searching of identifiers is automatic. But there are additional options available, accessed via the Search menu.

'Find' and 'Find again' match text in the active window.

'Extract from Window' copies matching text from the active window into a new window. 'Extract from file' does the same except that the source of input is a (disk) file, not a program window.

'Extract by Name' operates on the full Masterfile and automatically provides an exact match when you want to perform an ambiguous lookup. For example, 'Extract by Name''mitomycin' yields 22 matches of compounds whose name contains that string, instead of the single compound whose name is 'mitomycin'.

Because WLN records are present for almost all compounds in the 'Master' Thor database, a special kind of 'structural' searching is available. Both SMILES and WLN are string notations that describe compound structure, but unlike SMILES notation, WLN encodes structural groups directly. It is not necessary to learn the WLN to take advantage of this capability– the 'WLN dictionary' provides a concordance between WLN and many classes of structures.

Boolean Logic

ClogP supports simple Boolean logic in search queries; three operations are possible, via the reserved words 'and', 'or', & 'not'. Restated in English, these terms have the meaning:

and	=	both
or	=	either
not	=	opposite

Some useful combinations of these logical operations are illustrated below (set 'activity' as the default identifier before running these searches):

Syntax	Example	Result
A and B	poison and antidote	Require both terms
A or B	poison or antidote	Accept either term

A and not B   poison and not antidote   Match 'poison' but discard 'antidote'

For fastest operation, place restrictive queries first:

      bromo and benzoic acid   slower, because 'bromo' is more common
      benzoic acid and bromo   faster

In all cases, the end result is independent of how the query is entered.

Mixing and grouping (with parentheses) of 'and' with 'or' queries are not allowed. Chains of up to 9 and's & or's are allowed.

Shortcuts

ClogP supports a variety of keyboard shortcuts. Windows users should substitute the 'ctrl' key for the cmd- key used below.

Standard keyboard

Standard shortcuts for document and text manipulation:

cmd-N	new document
cmd-O	open document
cmd-W	close document
cmd-S	save document
cmd-Q	quit
cmd-C	copy
cmd-V	paste
cmd-X	cut
cmd-A	select all
cmd-F	find
cmd-G	find next
cmd-P	print
delete	same as cmd-X (only in 'ClogP output' window)

Extended keyboard

On the extended keyboard, these additional shortcuts are available:

page up	scroll up one page of text
page down	scroll down one page of text
home	move to first line of document
end	move to last line of document
HELP	same as cmd-H

Special to ClogP for Macintosh

The following keyboard shorcuts are special to the Macintosh version of ClogP:

cmd-R	Run one
ENTER	Run one
cmd-M	Run many
cmd-J	Run page
cmd-B	Run batch
cmd-1	Bring "SMILES input" window to front
cmd-2	Bring "Model output" window to front
cmd-3	Bring "Depiction" window to front