



# XLOGP3 v 3.2.0

## User Manual



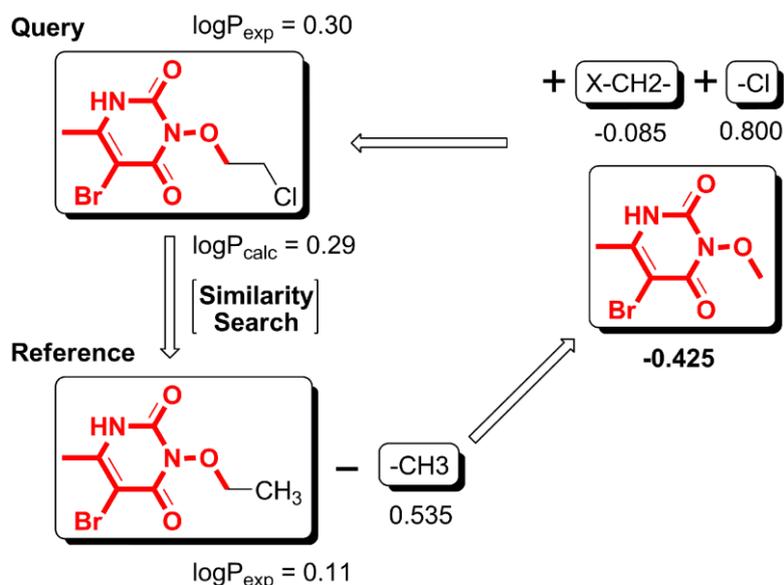
---

## Table of Contents

1. Overview on XLOGP3 .....	1
2. Program Installation.....	3
2.1 System Requirements.....	3
2.2 Installation.....	3
3. Program Usage.....	5
3.1 Getting Started .....	5
3.2 Screen a Compound Library .....	6
3.3 How to Use XLOGP3 .....	6
3.4 Process Multiple Files.....	7
3.5 Program Outputs .....	8
3.6 Warning and Error Messages.....	9
4. On Additional Knowledge Sets .....	10
4.1 Knowledge Set Preparation.....	10
4.2 Supply Additional Knowledge Sets.....	11
5. On-line Demo of XLOGP3.....	11
6. Appendix: Definition of Druglikeness Properties.....	12
Contact Information .....	13
References.....	13

## 1. OVERVIEW ON XLOGP3

The logarithm of the partition coefficient between *n*-octanol and water, also referred to as  $\log P$ , has been widely used in quantitative structure-activity relationship (QSAR) studies as a key parameter for characterizing lipophilicity. We have developed a new method, *i.e.* XLOGP3, for the fast calculation of  $\log P$ . Its predecessor, XLOGP and XLOGP2 ([ref. 1](#) and [2](#)), which are atom-additive methods with well-defined correction factors, have already gained their popularity in this field for their applicability and fair accuracy. As the latest release of this series, XLOGP3 has implemented an optimized atom typing scheme and is calibrated on a much larger training set. More importantly, based on the assumption that compounds with similar structures have similar properties, XLOGP3 introduces a new strategy by predicting the  $\log P$  value of a query compound based on the known  $\log P$  value of a structural analog, which is illustrated in Figure. 1. This new strategy improves the accuracy of XLOGP3 remarkably. It also provides users the possibility of utilizing their in-house  $\log P$  data. For more detailed information on XLOGP3, please refer to [ref. 3](#).



**Figure 1.** The basic calculation procedure of XLOGP3

The following two tables summarize the results of a comparison of XLOGP3 and some other popular  $\log P$  methods. The first test set consists of 406 FDA-approved drug molecules;

while the second one consists of 219 oligo-peptides.

**Table 1.** Results of Some log*P* Methods on 406 FDA Approved Small-Molecule Drugs

Models	$R^2$ <sup>a</sup>	RMSE <sup>b</sup>	MUE <sup>c</sup>
XLOGP3	0.872	0.72	0.51
XLOGP3-AA <sup>d</sup>	0.847	0.80	0.57
ALOGPS <sup>e</sup>	0.908	0.60	0.42
CLOGP	0.838	0.88	0.51
TOPKAT	0.815	0.88	0.56
AlogP98	0.802	0.90	0.64
XLOGP2	0.777	0.95	0.68
KOWWIN	0.771	1.10	0.63
HINTLOGP	0.491	1.93	1.30

<sup>a</sup> Correlation coefficients

<sup>b</sup> Root mean squared error between experimental and predicted values

<sup>c</sup> Mean unsigned error between experimental and predicted values

<sup>d</sup> XLOGP3 pure atom-additive model

<sup>e</sup> The test set is contained in the training set of ALOGPS

**Table 2.** Results of Some log*P* Methods on 219 Oligo-peptides

Models	$R^2$	RMSE	MUE
XLOGP3 <sup>a</sup>	0.932	0.32	0.24
XLOGP3 <sup>b</sup>	0.758	0.71	0.55
XLOGP3-AA	0.824	0.72	0.59
PLOGP	0.832	0.46	0.28
ALOGPS	0.765	0.73	0.54
TOPKAT	0.706	0.74	0.43
CLOGP	0.536	0.97	0.75
AlogP98	0.086	2.07	1.62
XLOGP2	0.078	2.24	1.78
KOWWIN	0.075	2.18	1.70
HINTLOGP	0.007	2.92	2.34

<sup>a</sup> When peptides are included in the knowledge set

<sup>b</sup> When peptides are not included in the knowledge set

The algorithm of XLOGP3 is developed by collaborations between Prof. Renxiao Wang's group at the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences and Prof. Luhua Lai's group at the College of Chemistry, Peking University. The full copyright of XLOGP3 is held by the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences.

## 2. PROGRAM INSTALLATION

### 2.1 System Requirements

XLOGP3 is a command line oriented program. It can be executed in a shell (e.g., C-shell, TC-shell, or B-shell on UNIX/Linux systems) or at a Windows or DOS command prompt. The following hardware platforms and operating systems are supported:

- ✧ x86 platforms (32-bit or 64-bit) running Linux<sup>®</sup>, kernel 2.4 or above
- ✧ x86 platforms running Microsoft<sup>®</sup> Windows<sup>®</sup> 98/2000/2003/XP

### 2.2 Installation

The XLOGP3 program is written in C++ and has been tested on UNIX/Linux and Windows platforms. It is provided as executable binary codes for each supported platform. To install the command line version of XLOGP3 please follow the instructions below.

1. Download the program package via the web server of Computational Chemical Biology Group (CCBG) at <http://www.sioc-ccbq.ac.cn/software/xlogp3/>. Please find the appropriate executable for your own platform.

Platform	Executable name
UNIX/Linux x86	<i>xlogp3.lnx.x86</i>
UNIX/Linux x86_64	<i>xlogp3.lnx.x86_64</i>
Microsoft Windows	<i>xlogp3.win32.exe</i>

2. Copy the downloaded package to a directory where XLOGP3 will be installed and uncompress it. A subdirectory named *xlogp3* will be created, *i.e.* the installation path of XLOGP3. It contains the following folders:

-  **bin/** contains executable file of XLOGP3
-  **manual/** contains user manual of XLOGP3
-  **example/** contains example files of XLOGP3
-  **parameter/** contains default parameter files of XLOGP3

For the sake of convenience, the installation path of XLOGP3 will be referred to as

*XLOGP3-Install-Dir* throughout the rest of this manual.

### 3. Environment variable configuration.

In order to execute XLOGP3 from any other folder or directory, we strongly recommend you to add the XLOGP3 executable and installation path to your system environment variables.

- a) For UNIX /Linux systems (x86/x86\_64 Linux)

If you are using a CSH or TCSH, please add the following contents to the file *.cshrc* under your home directory:

```
setenv XLOGP3_HOME XLOGP3-Install-Dir  
set path= ($path $XLOGP3_HOME/bin)
```

If you are using a BASH, please add the following contents to the file *.bashrc* or *.bash\_profile* under your home directory: (**Note:** there is no blank around '=')

```
export XLOGP3_HOME=XLOGP3-Install-Dir  
export PATH=$PATH:$XLOGP3_HOME/bin
```

If you are using other shells or have problem with installation, please contact to your system administrator for help.

- b) For Microsoft Windows Platforms (98/2000/2003/XP)

If you are using Windows 98 platform, please add the following contents to the file *Autoexec.bat* under your system installation path (typically in C:\)

```
set XLOGP3_HOME=XLOGP3-Install-Dir  
set PATH=%PATH%; XLOGP3-Install-Dir\bin
```

If you are using 32-bit platform (Windows 2000/2003/XP), open the user variables dialog via “**Control Panel → Performance and Maintenance → System → Advanced → Environment Variables**”. Create a new variable of XLOGP3\_HOME, set its value as *XLOGP3-Install-Dir*, then add *XLOGP3-Install-Dir*\bin to the user variable PATH (if not exist, create one).

### 3. PROGRAM USAGE

#### 3.1 Getting Started

The general synopsis for using XLOGP3 is:

**xlogp3 [-p[v]] *inputMoleFile* *outputResultFile* [exTTDB]**

XLOGP3 requires two necessary parameters, *i.e.* *inputMoleFile* and *outputResultFile*, and two optional parameters (*i.e.* *-pv* and *exTTDB*). Parameter *inputMoleFile* is the input filename containing molecules to be calculated. File extension of *inputMoleFile* is very important (but case-insensitive) because XLOGP3 will judge an input file format according to its file extension. The SYBYL MOL2 and MDL SDF/MOL/MDL/RDF formats are currently supported. *The MOL2 and SDF formats are recommended.* Parameter *outputResultFile* is the output filename containing calculation results. Parameter *-v* will give full information of program's output, otherwise the simplest. If you are dealing with a large database, turn off this parameter can cut down the output to minimum.

If optional parameter *-p* is provided, XLOGP3 will also calculate some druglikeness properties such as molecular weight, number of hydrogen bond donors and acceptors, number of rotatable bonds and number of rings (detailed information can be found in Section 6.1). Another optional parameter *exTTDB* indicates a user in-house knowledge set, if not provided, XLOGP3 will search only in default knowledge set (detailed information can be found in Section 4). A so-called file *xlogp3.log* will be created under current working directory to monitor the warning and error messages of XLOGP3 (detailed information can be found in Section 3.6).

#### 3.2 Screen a Compound Library

In practice, user may want to extract certain compounds from a large library for further study according to some restrictions, such as Lipinski's "rule of five". XLOGP3 provides user the filtering tool to screen a compound library using user-adjustable rules of the druglikeness properties. Just execute:

**xlogp3 -s *inputMoleFile* *outMoleFile* *ruleFile***

All parameters are indispensable. Parameter **-s** tells program to perform screen function, parameter ***inputMoleFile*** is the input filename of compound libraries to be screened. All compounds that satisfy the rules that defined in the file ***ruleFile*** will be written to a ***outMoleFile***. There is an example of ***ruleFile*** named ***default.rule*** under **example/** subdirectory in the installation path of XLOGP3 (shown in Figure 2).

```
#
# RANGE OF ACCEPTED MOLECULAR WEIGHT
MOLECULAR_WEIGHT    0    500
#
# RANGE OF ACCEPTED NO. OF HYDROGEN BOND DONOR
NUMBER_HB_DONOR    0    5
#
# RANGE OF ACCEPTED NO. OF HYDROGEN BOND ACCEPTOR
NUMBER_HB_ACCEPTOR 0    10
#
# RANGE OF ACCEPTED NO. OF TOTAL NITROGEN AND OXYGEN ATOM
NUMBER_HB_ATOM     0    15
#
# RANGE OF PARTITION COEFFICIENT
LOGP                -2    5
#
# RANGE OF ACCEPTED NO. OF ROTATABLE BOND
NUMBER_ROTOR       0    10
#
# RANGE OF ACCEPTED NO. OF RING NUMBER
NUMBER_RING        0    10
```

**Figure 2.** An Example of filtering rules

This file is designed as flexible as possible so that users can build their own rules very easily (**Note:** lines start with the character ‘#’ will be ignored by program, so the users can comment out one or more rules by adding a ‘#’ as the starting character of that lines).

### 3.3 How to Use XLOGP3

There are 3 files under **example/** subdirectory: ***test.mol2***, ***test.TTDB*** and ***default.rule***. File ***test.mol2*** contains 6 small molecules for testing. File ***test.TTDB*** is provided as additional user knowledge set which contains structural information and experimental  $\log P$  value of the molecules in ***test.mol2***. File ***default.rule*** defines filtering rules as shown in Figure 2. In the rest part of this section, we will show some examples of using XLOGP3 with these files. For your convenience, please copy them into your working directory.

Example 1) The following command will calculate the  $\log P$  values of the molecules in

*test.mol2* using **default.TTDB** (this file locates in **parameter/** subdirectory) as the default knowledge set. Results with detailed information are saved in *result\_df1.txt*.

```
xlogp3 -v test.mol2 result_df1.txt
```

Example 2) In addition to log*P* calculation, the following command will also calculate optional druglikeness properties such as molecular weight, number of hydrogen bond donors and acceptors, number of rotatable bonds and number of rings. Results are saved in *result\_df2.txt*.

```
xlogp3 -p test.mol2 result_df2.txt
```

Example 3) The following command will use *test.TTDB* as an additional knowledge set for log*P* calculation. Results are saved in *result\_df\_ex1.txt*.

```
xlogp3 test.mol2 result_df_ex1.txt test.TTDB
```

A more convenient way of utilizing user-provided knowledge sets is to make a copy of the database file (here *test.TTDB*) in **parameter/** subdirectory, then execute:

```
xlogp3 test.mol2 result_df_ex2.txt
```

You will see the result, *i.e.* *result\_df\_ex2.txt* is identical to that of *result\_df\_ex1.txt*. Actually, all the files under **parameter/** subdirectory with the extension of “TTDB” will be regarded by XLOGP3 as a knowledge set (detailed information can be found in Section 4).

Example 4) The following command will screen a compound library, *i.e.* *test.mol2*. All compounds that satisfy the restrictions defined in *example.rule* will be output as *test\_OK.mol2*.

```
xlogp3 -s test.mol2 test_OK.mol2 example.rule
```

### 3.4 Process Multiple Files

For the sake of convenience, XLOGP3 can handle several files under the same directory at one time by supporting wildcard. Each file can contain single or multiple molecules. The following command will calculate log*P* for files under *yourDIR* subdirectory (**Note:** only supported formats described in Section 3.1 will be processed, otherwise skipped):

```
xlogp3 “yourDIR/*.*” outputResultFile
```

Please note the couple of double quotation marks “ ” is necessary, they avoid the automatic expansion of wildcard by a UNIX/Linux shell or Windows/DOS command prompt. XLOGP3 regards all the characters between “ ” as a single parameter and expand it internally if it is wildcard-contained.

### 3.5 Program Outputs

The verbose output (with `-v` parameter) of XLOGP3 is like the following:

```

Optional Druglikeness Properties of MOL HEaSC00584
-----
                                Molecular Weight: 132.1
                        No. of Hydrogen Bond Donors: 3
                No. of Hydrogen Bond Acceptors: 3
                        No. of Rotatable Bonds: 5
                No. of Nitrogen and Oxygen Atoms: 5
                                No. of Rings: 0

                                XLOGP3 of MOL HEaSC00584: -2.92
                with Similarity Index to Reference: 1.00
-----
+ Contribution from Reference's Experimental LogP: -2.92
- Contribution from Reference's Calculated LogP: -4.05
  + Contribution from Each Atom:
    1                                C.2.(=X).X: -0.760
    2                                N.am.h: -0.272
    3                                C.3.2h.X.pi: -0.264
    4                                O.2.(=C): 0.641
    5                                O.3.h.pi: -0.007
    6                                C.2.(=X).X: -0.760
    7                                N.3.2h: -0.730
    8                                C.3.2h.X.pi: -0.264
    9                                O.2.(=C): 0.641
  + Contribution from 'Amino Acid' Correction: -2.280

```

**Figure 3.** An Example of output of XLOGP3

Figure 3 lists the full output of one example molecule, including molecular name, optional druglikeness properties, final calculated  $\log P$ , similarity index to reference molecule and individual contribution. A similarity index cutoff 0.50 is currently applied in XLOGP3 for finding an appropriate reference as starting point. If the structural analog with the highest similarity index to given molecule that found in knowledge set is still below 0.50, it will be calculated with pure atom-additive model, *i.e.* XLOGP3-AA. Without `-v`, the output will be the simplest.

### 3.6 Warning and Error Messages

All the warning and error messages will be logged to a file named *xlogp3.log* and also written to screen simultaneously. We recommend XLOGP3 users to forward this file to us when problems are encountered with using XLOGP3. Table 3 lists some of the general warning and error messages.

**Table 3.** Common Warning and Error Messages of XLOGP3

Warning or Error Message	Description
Oops: syntax error	The general synopsis for using XLOGP3 is error. See Section 3.1 for detail.
Unsupported format X	Format X is not supported. The only supported formats are MOL2, SDF, RDF, MOL and MDL, and MOL2 and SDF are recommended.
Unable to open file X	Unable to read file X. It is normally because the file does not exist or the path provided is not correct.
Unable to find file matches pattern X	Unable to find file matches pattern X. See Section 3.4 for detail.
No TTDB found or supplied, XLOGP3 works in XLOGP3-AA mode	This may be because the default knowledge set ( <b>default.TTDB</b> ) is missing or the environment variables settings are not correctly configured. See Section 2.2 for detail.
Missing parameter for ATOM: X	The type of atom X is not defined in XLOGP3, we make a guess on its coefficient.
Unable to construct TT for X, skipped	Unable to construct Topological Torsion (TT) descriptor for molecule X, it will be skipped when building user knowledge set. A TT descriptor requires at least four consecutive non-hydrogen atoms.
Unable to get experimental logP value for X, skipped	Unable to get the experimental logP value for molecule X, it will be skipped when building user knowledge set. See Section 4.1 for detail.

## 4. ON ADDITIONAL KNOWLEDGE SET

Some researchers may maintain their in-house collections of  $\log P$  data. For understandable reasons, such data are not always available to the public. Conventional additive methods are normally provided to users as is and lack the ability of utilizing users' in-house data. We provide auxiliary tools in the XLOGP3 package so that the users may process their own data sets into the format accepted by XLOGP3 and supply them as the knowledge set used by XLOGP3. The importance of this feature should not be underestimated since in-house  $\log P$  data are perhaps many times more than publicly available data. We have already provided a default knowledge set, *i.e.* *default.TTDB*, under **parameter/** subdirectory. It contains all the compounds in the training set and two test sets of XLOGP3, ~9000 in total.

### 4.1 Knowledge Set Preparation

In order to process user data sets into the format accepted by XLOGP3, please execute:

```
xlogp3 -m inputMoleFile outTTDB nameLogpFile
```

This command will process all the molecules in *inputMoleFile* into a knowledge set named *outTTDB* (an extension of **TTDB** is recommended, *e.g.* *yourOwn.TTDB*). Parameter *nameLogpFile* is an index file containing the name and experimental  $\log P$  of molecule. It has the following format:

Name_of_Molecule_1	Experimental_logP_of_Molecule_1
Name_of_Molecule_2	Experimental_logP_of_Molecule_2
...	...

Molecular name must correspond to experimental  $\log P$  strictly because XLOGP3 will use molecular name as index for experimental  $\log P$  value (**Note:** molecular names should contain no blank). Some compounds may be skipped in the preparation step, see Table 3 for detail. Wildcard is also supported. The following command regards all the MOL2 files under *yourDIR* subdirectory as user data sets and tries to process them into *outTTDB*.

**xlogp3 -m "yourDIR/\*.mol2" outTTDB nameLogpFile**

## 4.2 Supply Additional Knowledge Sets

Once a knowledge set is ready for use, one may supply it as a parameter as following:

**xlogp3 inputMoleFile outputResultFile yourTTDB**

Or in a simpler form, one may put a copy under **parameter/** subdirectory then execute with more brief command:

**xlogp3 inputMoleFile outputResultFile**

Actually, all the files under **parameter/** subdirectory with the extension of "TTDB" will be regarded by XLOGP3 as knowledge set.

## 5. ON-LINE DEMO OF XLOGP3

**Figure 4.** On-line Demo of XLOGP3

XLOGP3 is available for on-line testing via the web server of Computational Chemical

Biology Group (CCBG) at <http://www.sioc-ccbq.ac.cn/software/xlogp3/>. User can submit a job through web (Figure 4) by uploading molecules in either of MOL2, SDF, RDF, MOL or MDL formats.

In addition to  $\log P$  calculation, some optional druglikeness properties such as molecular weight, number of hydrogen bond donors and acceptors, number of rotatable bonds and *etc.* Results can be downloaded once the job is finished (Figure 5).

Files Uploaded	
File Name	File Size
test.sdf	18.23 KB
XLOGP3 Result	
File Name	File Size
<a href="#">xlogp3_6993.logp</a>	11.22 KB
<a href="#">xlogp3_6993.err</a>	0 B

**Figure 5.** Results of an Job Submitted On-line.

## 6. APPENDIX: Definition of Druglikeness Properties in XLOGP3

- ✧ **Hydrogen Bond Donor:**  $sp^3$  hybridized nitrogen or oxygen atom with at least one hydrogen atom, *e.g.* -OH, -NH-, -NH<sub>2</sub>, -NH<sub>3</sub><sup>+</sup>.
- ✧ **Hydrogen Bond Acceptor:**  $sp^2$  hybridized oxygen atom in a carbonyl or nitro group;  $sp^3$  hybridized oxygen atom in hydroxyl group;  $sp^2$  nitrogen atom in an aromatic ring;  $sp^1$  nitrogen atom in cyano group.
- ✧ **Rotatable Bond:** must satisfy: a) single bond, b) not a ring bond, c) not a terminal bond except that connected to -OH, d) not connected to -CH<sub>3</sub>, -CX<sub>3</sub> (X = halogen atom), -C#N, -NH<sub>2</sub>, -NH<sub>3</sub><sup>+</sup>, -NO<sub>2</sub>, -PO<sub>3</sub><sup>2-</sup>, -CO<sub>2</sub><sup>-</sup>, -SO<sub>3</sub><sup>-</sup>, -*t*-Butyl, e) count only once if a triple bond has two rotatable bonds

## CONTACT INFORMATION

You are welcome to post your questions or any comments on XLOGP3 on an on-line discussion forum at <http://www.sioc-ccbg.ac.cn/forum/>. Or, you may contact us at the following address.

Prof. Renxiao Wang

*State Key Lab of Bioorganic and Natural Products Chemistry*

*Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences.*

*354 Fenglin Road, Shanghai 200032, China, P. R.*

*E-mail: [wangrx@mail.sioc.ac.cn](mailto:wangrx@mail.sioc.ac.cn)*

---

## REFERENCES

Users are required to cite the usage of XLOGP3, XLOGP2, or XLOGP in any resulting publication by making reference to the publications below.

1. Wang, R.; Fu, Y.; Lai, L. A New Atom-Additive Method for Calculating Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615-621. [XLOGP]
2. Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-additive Method. *Perspect. Drug Discov. Des.* **2000**, *19*, 47-66. [XLOGP2]
3. Tiejun Cheng, Yuan Zhao, Xun Li, Fu Lin, Yong Xu, Xinglong Zhang, Yan Li and Renxiao Wang\*, Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, *47* (in press) [XLOGP3]