



XLOGS软件

v.1

用户使用手册



目 录

1. XLOGS 程序简介	1
2. XLOGS 程序安装说明	2
2.1 系统需求.....	2
2.2 安装.....	2
3. 使用 XLOGS 程序	4
3.1 通用命令.....	4
3.2 过滤化合物库.....	4
3.3 XLOGS 程序使用示例	5
3.4 一次处理多个文件	6
3.5 输出结果说明	6
3.6 警告和错误	7
4. 用户自定义数据库.....	9
4.1 准备数据库	9
4.2 使用数据库	10
5. 附录.....	11
5.1 XLOGS 程序中类药性参数定义	11
联系信息.....	11

1. XLOGS 程序简介

XLOGS程序作为药物发现和设计过程中的辅助工具,能够快速准确地预测未知化合物的特性水溶解度(通常表示为 $\log S$)。XLOGS程序基于原子加和法原理,同时引入了分子模拟领域广泛流行的“最近邻”方法的思想。对于一个 $\log S$ 未知的化合物,XLOGS程序首先从数据库(其中包含已知化合物的结构和 $\log S$ 实验值等信息)中找到与给定分子结构最相近的类似物作为最佳参照,然后从该参照物的 $\log S$ 实验值出发加上二者结构差异的贡献值得出未知化合物的预测值。下图的例子简单说明了XLOGS程序的预测流程。

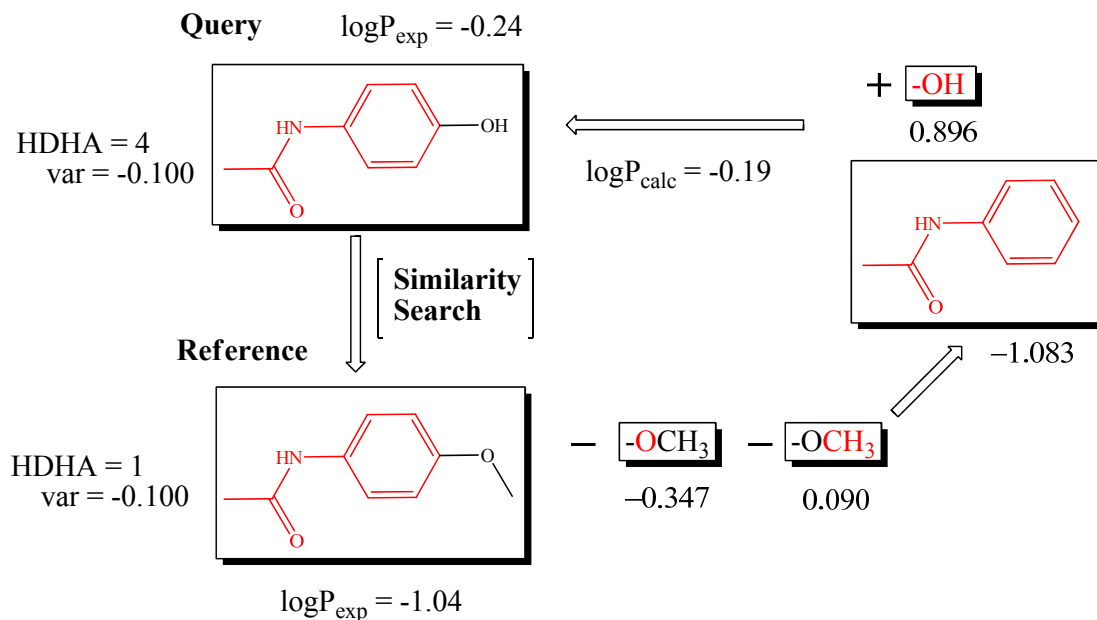


图 1 XLOGS 程序计算方法示例

XLOGS 程序的算法由中国科学院上海有机化学研究所王任小研究员课题组开发。

2. XLOGS 程序安装说明

2.1 XLOGS 系统需求

XLOGS程序是一个基于命令行的控制台应用程序，可以通过Unix/Linux系统下的shell (如csh, tcsh或bash) 以及Microsoft Windows系统命令行或DOS窗口执行。XLOGS程序可以运行于以下环境：

✧ Linux[®] 内核2.4或更高 (需要gcc 3.2或更高)

✧ Microsoft[®] Windows[®] 98/2000/2003/XP

(注：gcc为免费开源编译器集合，可在<http://gcc.gnu.org/>自由下载)





2.2 安装

XLOGS 程序采用标准 C++编写，并已在 UNIX/Linux 和 Windows 平台编译通过。请按照以下说明进行安装。

1. 请从 <http://www.sioc-ccbg.ac.cn/software/xlogs/>下载程序安装包，根据自己的系统环境选择合适的可执行文件。

系统平台	可执行程序名称
UNIX/Linux(32-bit)	xlogs_32
UNIX/Linux(64-bit)	xlogs_64
Microsoft Windows	xlogs.win32.exe

2. 将下载的程序安装包复制到用户所希望的XLOGS的安装位置，解压。这将生成xlogs子目录，即为XLOGS程序的安装目录，以下均简称作**XLOGS-Install-Dir**，其目录结构如下：

 bin/	包含XLOGS程序可执行文件
 manual/	包含XLOGS程序用户使用手册
 example/	包含XLOGS程序供练习示例文件
 parameter/	包含XLOGS程序默认参数文件

3. 设置环境变量

为了能够在任意位置调用XLOGS程序，我们建议将XLOGS程序的相关路径添加至用户系统环境变量中。

a) UNIX/Linux系统

如果用户使用csh（tcsh的设置相同），请编辑（或新建）用户主目录下的.cshrc文件，添加以下内容：

```
setenv XLOGS_HOME XLOGS-Install-Dir
```

```
set path= ($path $XLOGS_HOME/bin)
```

如果用户使用bash，请编辑（或新建）用户主目录下的.bashrc或.bash_profile文件，添加以下内容：（注意：等号两边没有空白）

```
export XLOGS_HOME=XLOGS-Install-Dir
```

```
export PATH=$PATH:$XLOGS_HOME/bin
```

如果用户使用其它shell，请参考上面的设置修改相应的配置文件。

b) Windows系统（98/2000/2003/XP）

如果用户使用Windows 98系统，请编辑（或新建）系统（通常为C盘）根目录下的Autoexec.bat文件，添加以下内容：

```
set XLOGS_HOME=XLOGS-Install-Dir
```

```
set PATH=%PATH%; XLOGS-Install-Dir\bin
```

如果用户使用Windows XP系统（Windows 2000，Windows 2003设置类似），请依次打开“控制面板→性能和维护→系统→高级→环境变量”或者“我的电脑→（右键）属性→高级→环境变量”，调出环境变量设置窗口。在用户变量窗口中新建名为XLOGS_HOME的环境变量，值为*XLOGS-Install-Dir*，然后将*XLOGS-Install-Dir*\bin添加至（或新建）用户变量的PATH变量中（注意：不同路径之间使用分号隔开）。

3. 使用 XLOGS 程序

3.1 通用命令

使用XLOGS程序计算分子logS的通用命令为：

xlogs [-p[-v]] inMoleFile outResultFile [exTTDB]

XLOGS 程序接受两个必要参数（inMoleFile, outResultFile）和两个可选参数（-p, exTTDB）。其中 inMoleFile 为包含需要计算 logS 的分子的输入文件名，文件名后缀非常重要（但不区分大小写），因为 XLOGS 程序将自动根据所提供文件名的后缀来识别输入文件的格式。目前支持的文件格式有 SYBYL MOL2 格式以及 MDL 系列 SDF、MOL、MDL 和 RDF 格式。为了达到最佳计算结果，建议用户使用 MOL2 或 SDF 作为输入文件首选格式。outResultFile 文件用来保存计算结果。参数-v 将会列出程序详细的输出信息，建议在处理大型数据库时关闭该选项。如果打开可选参数-p，XLOGS 程序除计算 logS 之外，还将对分子的类药性参数（如分子量、氢键给体数目、氢键受体数目、可旋转键数目以及环的数目等，有关 XLOGS 程序类药性参数详细说明见手册 5.1）进行计算。可选参数 exTTDB 为用户外挂数据库，如果未提供该参数，程序将在默认数据库中进行搜索。此外，程序在当前工作目录下生成名为 XLOGS.log 的日志文件以记录 XLOGS 程序运行情况，包括所有的警告和错误信息。

3.2 过滤化合物库

在实际应用中，用户可能需要从一个大型的化合物库中选择一些符合要求的分子作后续研究。XLOGS 程序可以通过分析化合物库中分子的类药性参数，按照用户预设的约束条件，将那些不满足约束条件的分子过滤掉。命令如下：

xlogs -s inMoleFile outMoleFile ruleFile

参数-s 告知程序将执行过滤化合物库功能，输入文件 inMoleFile 包含需要进行过滤的分子，满足约束条件的分子存储在 outMoleFile 中，而用户设定的约束条件由 ruleFile 定义。在 [example/](#)目录下有一个名为 default.rule 的示例文件，该文件内容如图 2 所示：

```

### DEFINE YOUR OWN RULE BELOW ###
### COMMENT ONE LINE WILL IGNORE TAHT RULE ###
#
# 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

```

图 2 自定义筛选约束条件

用户可根据自己的实际需要修改其中的内容来自定义约束条件，如定义分子量的范围从 200-500，环的个数从 3-5 等。通过在行首添加字符#可以屏蔽对应的约束条件，方便用户灵活使用。

3.3 XLOGS 使用示例

在 [example/](#)目录下包含 3 个文件：test.sdf、test.TTDB 和 default.rule。test.sdf 文件包含有 7 个测试分子。test.TTDB 作为用户提供的自定义数据库文件，记录了 test.sdf 中所有分子的结构和 logS 实验值等信息。default.rule 文件定义了一些分子类药性参数的约束条件，用于过滤 test.sdf 文件中的分子。下面我们将以它们为例来详细介绍如何使用 XLOGS 程序。方便起见，请将这些文件复制到用户的工作目录下。（注意：请确保已经按照上面的说明设置好相应的环境变量）

1) 执行以下命令，XLOGS 程序将在默认数据库 default.TTDB（该文件在 [parameter/](#)目录下）来搜索与 test.sdf 中分子相似度最高的分子作为参照，计算结果存储在 result_df1.txt 中。

xlogs test.sdf result_df1.txt

2) 下面的命令除计算 test.sdf 中所有分子的 logS 外，还将计算诸如分子量、氢键给

体、受体数目、可旋转键数和环个数等类药性参数。

```
xlogs -p test.sdf result_df2.txt
```

3) 执行以下命令，XLOGS 程序将同时在默认数据库 default.TTDB 和用户提供的自定义数据库 test.TTDB 中进行搜索，计算结果存储在 result_df_ex1.txt 中。

```
xlogs test.sdf result_df_ex1.txt test.TTDB
```

用户也可以先将自定义数据库 test.TTDB 复制到 [parameter/](#) 目录，然后，执行命令

```
xlogs test.sdf result_df_ex2.txt
```

可以看到 result_df_ex1.txt 和 result_df_ex2.txt 的结果完全一致，但命令更为简洁。事实上，[parameter/](#) 目录下所有以 TTDB 为后缀的文件都将作为搜索的目标数据库。

4) 执行以下命令，XLOGS 程序会对 test.sdf 中所有分子根据文件 default.rule 中定义的分子类药性参数的约束条件进行过滤，满足条件的分子将存储在 test_new.sdf 文件中。

```
xlogs -s test.sdf test_new.sdf example.rule
```

3.4 一次处理多个文件

在上面的示例中，输入文件只有一个。为方便用户使用，XLOGS 程序能够处理同一目录下包含多个输入文件（以通配符的方式支持）的情况。每个输入文件又可以是包含单个和多个分子，XLOGS 程序能够自动识别。下面的命令将计算 yourDIR 目录下所有的 MOL2 格式文件中的所有分子的 logS。

```
xlogs "yourDIR/*.mol2" outResultFile
```

请注意命令中的一对双引号“”，它们是必须的，这是为了避免 Linux shell 以及 Window 命令行或 DOS 窗口对通配符自动展开。XLOGS 程序将包含在一对双引号“ ”内所有字符作为单一参数，如果该参数包含通配符，则在程序内部对其进行展开。如果希望对 yourDIR 目录所有以 a 开头文件中的分子进行过滤，可以执行（注意：XLOGS 程序只处理手册 3.1 节提到的文件格式，而不被支持的格式将被跳过）：

```
xlogs -s "yourDIR/a*.*" outResultFile ruleFile
```

3.5 输出结果说明

XLOGS 程序的一般输出结果如图 3 所示。


```

Optional Druglikeness Properties of MOL Acetaminophen
-----
Molecular Weight: 151.2
No. of Hydrogen Bond Donors: 2
No. of Hydrogen Bond Acceptors: 2
No. of Rotatable Bonds: 3
No. of Nitrogen and Oxygen Atoms: 3
No. of Rings: 1

XLOGS of MOL Acetaminophen: -0.24
with Similarity Index to Reference: 1.00
Contribution from Reference's Experimental LogS: -0.24
Contribution from Reference's Calculated LogS: 0.19
+ Contribution from Each Atom:
1          O.3.h.pi: 0.896
2          O.2.(=C): -0.038
3          N.am.h: 0.509
4          C.ar.(-X): -0.118
5          C.ar.h: -0.248
6          C.ar.h: -0.248
7          C.ar.h: -0.248
8          C.ar.h: -0.248
9          C.ar.(-X): -0.118
10         C.2.(=X).X: 0.196
11         C.3.3h.X.pi: 0.247
+ Contribution from 'Hydrogen Donor and Acceptor' Correction: -0.398

```

图 3 XLOGS 的输出结果示例

图中列出了其中一个分子的计算结果，包括分子的名称、可选的类药性参数（分子量、氢键给体数目、氢键受体数目、氮氧原子总数、可旋转键数目和环个数）、最终的 logS 计算值、与所选参照分子的相似度系数及每一项对最终 logS 的贡献值（包括最佳参照对象的实验 logS 值、最佳参照对象的计算 logS 值、每一个原子或校正项的贡献值）。在 XLOGS 程序中为了在一定程度上保证计算结果的准确性，设定所选参照分子与给定分子间最低的相似系数为 0.5。如果数据库搜索所得最佳参照分子与给定分子间的相似系数仍低于该数值，XLOGS 程序将采用纯粹的原子加和法进行计算，即 XLOGS-AA。因此，相似度系数可以在一定程度上说明计算结果的可靠性。值越大，计算所选的参照对象与给定分子越相似，结果也越可靠。特别地，如果在数据库中找到的最佳参照分子与给定分子相似度为 1.00，说明给定分子已经在数据库中存在，此时的 logS 计算值实际上就是该分子的 logS 实验值。

3.6 警告和错误

XLOGS 程序在运行过程中可能遇到的问题将会记录在名为 XLOGS.log 的日志文件

中，相关警告或出错信息也将在屏幕同步输出。通过该文件有助于分析程序非正常运行的原因，建议用户在遇到问题时将该文件通过 email 反馈给我们。表 1 中列出了一些 XLOGS 程序常见的警告或出错信息。

表 1 XLOGS 程序常见警告或出错信息

警告或出错信息	说明
Ooops: syntax error	XLOGS 程序使用语法出错。使用一般语法详见手册 3.1
Unsupported format <i>kkk</i>	XLOGS 程序不支持文件 <i>kkk</i> 的格式，文件被跳过。当前仅支持 MOL2、SDF、RDF、MOL、MDL 文件格式
Unable to open file <i>nnn</i>	XLOGS 程序无法打开文件 <i>nnn</i> ，通常因为文件不存在或提供的路径不正确
Unable to find file matches pattern <i>xxx</i>	XLOGS 程序没有找到匹配模式为 <i>xxx</i> 的文件。详见手册 3.4
No TTDB found or supplied, XLOGS works in XLOGS-AA mode	XLOGS 程序没有找到用于搜索的数据库文件，将采用原子加和法计算 logS。这可能因为系统环境变量设置不正确或者默认的数据库文件（default.TTDB）丢失
Unable to construct TT for <i>yyy</i> , skipped	XLOGS 程序构建 Topological Torsion 描述符失败。分子 <i>yyy</i> 太小，不满足 Topological Torsion 描述符的至少需要连续四个非氢原子的要求，该分子在构建用户数据库时被跳过
Unable to get experimental logS for <i>zzz</i> , skipped	XLOGS 程序无法找到分子 <i>zzz</i> 的 logS 实验数据，该分子在构建用户数据库时被跳过 (注：构建用户数据库需要提供分子名称与 logS 实验数据的索引文件，详见手册 4.1)

4. 用户自定义数据库

实际的应用中，用户可能拥有自己的化合物库及相关的 logS 实验数据，但就我们所知，在现有的 logS 预测程序中，几乎没有哪个程序能够充分利用甚至关注过这部分资源，而 XLOGS 程序则允许以外挂数据库的方式对这部分资源加以利用。在 [parameter/](#) 目录下，我们已经提供了一个名为 default.TTDB 数据库，其中包含 4129 个小分子的结构及其 logS 实验数据的信息，在默认情况下它将作为搜索给定分子的最佳参照对象的目标数据库。如果用户能够提供自己的数据库，XLOGS 程序可将它与默认数据库 default.TTDB 一起作为搜索的目标数据库，扩大搜索范围和提高命中率。

4.1 准备数据库

为了将用户数据转换成 XLOGS 程序可以识别的数据库格式，我们需要利用到 XLOGS 程序的数据库准备功能，使用方法如下：

xlogs -m inMoleFile outTTDB nameLogSFile

其中参数 -m 告知程序将执行数据库准备功能，参数 inMoleFile 为包含分子结构的输入文件名，可接受的文件格式同上。参数 outTTDB 为数据库输出文件名（推荐以 TTDB 作为后缀名，例如 your_own.TTDB），用于保存成功转换的数据。参数 nameLogSFile 为包含分子名称与分子 logS 实验值的索引文件，XLOGS 程序需要根据分子名称从该索引文件找到给定分子的 logS 实验值，并将其输出到 outTTDB 文件中。nameLogSFile 文件格式如下：

分子 1 的名称	分子 1 的 logS 实验值
分子 2 的名称	分子 2 的 logS 实验值
...	...

其中分子名称与实验 logS 需要严格一一对应（注意：分子名称中不能包含空白）。

对于特别小或者在 nameLogSFile 文件中找不到对应 logS 实验值的分子，将在构建数据库时被跳过，同时 XLOGS 程序将给出相应的警告信息（详见手册 3.6）。

数据库准备功能同样也支持通配符，能够同时处理同一目录下的多个输入文件。如：

xlogs -m “yourDIR/*.mol2” outTTDB nameLogSFile

也可以执行

xlogs -m “yourDIR/a*.*” outTTDB nameLogSFile

（注意：命令中一对引号“”是必须的）

4.2 使用数据库

在准备好数据库之后，只要将其作为 XLOGS 程序的可选参数即可使用，如：

```
xlogs inMoleFile outResultFile yourTTDB
```

或者

```
xlogs -p inMoleFile outResultFile yourTTDB
```

更简单地，用户可以将自己的数据库文件复制一份到 **parameter/** 目录，然后通过以下更为简洁的命令使用数据库：

```
xlogs inMoleFile outResultFile
```

或者

```
xlogs -p inMoleFile outResultFile
```

事实上，**parameter/** 目录下所有以 TTDB 为后缀名的文件都将作为搜索的目标数据库。因此，用户可以自由地向其中添加或删除数据库文件来调节目标搜索库的大小。（注意：XLOGS 程序所使用数据库为具有特殊格式的二进制文件，请不要将其它格式的文件以 TTDB 为后缀存放在 **parameter/** 目录下，否则可能引起程序未知错误）

5. 附录

5.1 XLOGS 程序中类药性参数定义

- ✧ **氢键给体**: 连接有 H 原子的 N 或 O 原子, 如-OH, -NH-, -NH₂, -NH₃⁺等
- ✧ **氢键受体**: sp² 杂化的 O, 如羰基或硝基上的 O 原子; 羟基上的 O 原子; 芳香环上连接度为 2 的 N 原子, 如吡啶环上 N 原子; 氰基上的 N 原子
- ✧ **可旋转键**: 需满足以下要求。a) 必须是单键; b) 环上键不是可旋转键; c) 末端键不是可旋转键; d) 连接-OH 的键为可旋转键; e) 连接-CH₃, -CX₃ (X 为卤素原子), -C#N, -NH₂, -NH₃⁺, -NO₂, -PO₃²⁻, -CO₂⁻, -SO₃⁻, -t-Butyl 等不是可旋转键; f) 如果叁键两边各连接一个可旋转键则只算一个。

联系信息

用户可以通过以下地址和我们取得联系:

王任小 研究员

中国科学院上海有机化学研究所

生命有机化学国家重点实验室

上海市徐汇区枫林路 354 号, 邮编 200032

E-mail: wangrx@mail.sioc.ac.cn

为了更好地改进我们的程序, 欢迎大家多提宝贵意见和建议。